A MORPHING-BASED TECHNIQUE FOR THE VERIFICATION OF DETERMINISTIC AND ENSEMBLE PRECIPITATION FORECASTS

A Dissertation

by

FAN HAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Istvan Szunyogh
Committee Members,	Ping Chang
	Ramalingam Saravanan
	Mikyoung Jun
Head of Department,	Ping Yang

May 2017

Major Subject: Atmospheric Sciences

ABSTRACT

A new morphing based technique is proposed for the verification of deterministic precipitation forecasts. It provides accurate estimates of the components of the precipitation forecast error for a significantly longer range of the location error, and it is computationally more efficient than its earlier version. The behavior of the technique is illustrated by its application to idealized and realistic deterministic forecast examples.

A set of diagnostic equations that take advantage of the new technique is introduced for the verification of ensemble forecasts of the precipitation. The diagnostic equations are applied to operational ensemble forecasts of the named winter storms of the United States from the 2014/2015 and 2015/2016 winter storm season. The results indicate that the forecasts can better predict the uncertainty in the location of the storms than the uncertainty in the amount and the structure of the precipitation that they produce.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Istvan Szunyogh for sharing his wisdom on the development of my dissertation research. This dissertation would have not been possible without his great support, constant encouragement, and tremendous patience in discussing research results with me and revising our manuscripts over and over again. My appreciation also goes to my committee members, Dr. Ping Chang, Dr. Ramalingam Saravanan, and Dr. Mikyoung Jun, for providing valuable and insightful suggestions on my research and the dissertation.

I also would like to thank everyone in Dr. Szunyogh's research group and all the fellow graduate students in the Department of Atmospheric Sciences for providing an extremely friendly working environment.

My final and special thanks go to my family and my best friend, Dr. Guanglin Tang, who means so much to me in every aspect of my life. Thank you so much, Guanglin, for always being there for me.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Istvan Szunyogh [advisor], Professor Ping Chang, Professor Ramalingam Saravanan of the Department of Atmospheric Sciences, and Professor Mikyoung Jun of the Department of Statistics.

All work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projection Program, Grant NA16OAR4310082.

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGMENTS iii

CONTRIBUTORS AND FUNDING SOURCES iv

TABLE OF CONTENTS v

LIST OF FIGURES vii

LIST OF TABLES ix

1. INTRODUCTION 1

2. PRECIPITATION VERIFICATION METHODS 4

2.1 Spatial Verification Methods 4
2.1.1 Introduction 4
2.1.2 Object-based Verification 5
2.1.3 Morphing-based Verification 5
2.1.4 Applications 6
2.2 Ensemble Verification 6
2.2.1 Introduction 6
2.2.2 Rank Histogram 7
2.2.3 Continuous Rank Probability Score (CRPS) 7

3. THE PROPOSED VERIFICATION TECHNIQUE 9

3.1 The Original Morphing Technique 9
3.2 The Modified Morphing Technique 10
3.3 ASSIM 11
3.4 Illustration of the Verification Technique 14
3.4.1 Idealized Example 14
3.4.2 ICP Cases 15
3.4.2.1 Idealized Cases 15
3.4.2.2 Realistic Cases 18
3.4.3 Winter Storms 20

4. METRICS OF UNCERTAINTY FOR THE VERIFICATION OF ENSEMBLE PRECIPITATION FORECASTS 24

4.1 Finding a Matching Forecast System 24
4.2 Position Forecasts 25
4.2.1 Definition of the Position Uncertainty 25
4.2.2 Position Bias 28
4.2.3 Position Spread 28
4.3 Amplitude Forecasts 30
4.4 Structure Forecasts 33

5. FORECAST AND VERIFICATION DATA 35

5.1 Winter Storm Case Description 35
5.2 Forecast Data 35
5.3 Verification Data 35

6. VERIFICATION RESULTS AND DISCUSSIONS 41

6.1 Sensitivity of the Average K' to the Choice of the ASSIM Criteria 41
6.2 Uncertainty Results for $\delta = 0.8$ 43
6.2.1 Location Uncertainty 43
6.2.2 Amplitude Uncertainty 43
6.2.3 Structure Uncertainty 43
6.2.3.1 Variance 43
6.2.3.2 Correlation 45
6.3 Sensitivity of the Uncertainty Results to the Choice of the ASSIM Criteria 45

7. CONCLUSIONS 53

REFERENCES 56

LIST OF FIGURES

FIGURE Page

3.1 Illustration of the steps of the proposed morphing technique for an idealized case in which the forecast has only location error (see Section 3.4.1 for details). 16

3.2 The estimates of the displacement errors for the series of simulated forecasts described in Section 3.4.1 that were obtained by (black) the technique of HS16 and (red) the proposed technique of the present paper. 17

3.3 Illustration of a realistic ICP example. Shown are (left) an 24 h forecast of the 1 h precipitation total by the NCEP WRF model, (middle) the related Stage II analysis of 1 h accumulation, and (right) the shifted forecast obtained by the proposed morphing technique. The black arrow in the right panel indicates the magnitude and direction of $d\mathbf{X}_{mean}$ (the total shift vector). 19

3.4 Illustration of the example of Winter Storm Cato. (left) 240 h NCEP forecast of the 6 h precipitation total at 0000 UTC November 26, 2014, (middle) the verifying analysis, and (right) the forecast shifted by $d\mathbf{X}_{mean}$. The grey shading in the left and right panel indicates the contour of the precipitation field in the verifying analysis (middle panel). The black arrows in the right panel are the shift vectors of the iterations of the morphing algorithm. 21

3.5 Same as Fig 3.4, except for the 120 h forecast of winter storm Sparta at 0000 UTC, March 05, 2015. 21

3.6 Same as Fig 3.4, except for the 360 h forecast of winter storm Sparta at 0000 UTC, March 05, 2015. 23

5.1 (left) ECMWF 6 h accumulated precipitation forecast, (middle) regridded Stage IV 6 h precipitation analysis, and (right) combined ST4-ECMWF precipitation field for Winter Storm Frona starting at 0000 UTC, December 29, 2014. 39

5.2 Same as Fig. 5.1, except for Winter Storm Quo starting at 0000 UTC, March 02, 2016. 39

5.3 Histogram of the percentages of the Stage IV total precipitation over the total precipitation of the composite ST4-ECMWF precipitation of the 133 samples. 40

6.1 Evolutions of the average K' and percentages of ensembles in which $K' \geq 2$ with different δ values: (purple) $\delta = 0.6$, (blue) $\delta = 0.7$, (green) $\delta = 0.8$, and (red) $\delta = 0.9$. 42

6.2 Evolution of σ_{pos} (ensemble spread) and δ_{pos} (ensemble mean error). . . . 44

6.3 Evolution of σ_{amp} (ensemble spread) and δ_{amp} (ensemble mean error). . . 44

6.4 Evolution of v_f and v_a. 46

6.5 Evolution of c_f and c_a. 46

6.6 Evolutions of σ_{pos} and δ_{pos} for $\delta = 0.6, 0.7, 0.8$, and 0.9. 48

6.7 Evolutions of σ_{amp} and δ_{amp} for $\delta = 0.6, 0.7, 0.8$, and 0.9. 49

6.8 Evolutions of v_f and v_a for $\delta = 0.6, 0.7, 0.8$, and 0.9. 50

6.9 Evolutions of $v_a - v_f$ for $\delta = 0.6, 0.7, 0.8$, and 0.9. 51

6.10 Evolutions of c_f and c_a for $\delta = 0.6, 0.7, 0.8$, and 0.9. 52

LIST OF TABLES

TABLE Page

3.1 Summary of the results for the idealized ICP cases. 16

3.2 The correlation between the different objective measures and the subjective expert score for the realistic ICP forecasts. We use the values of the expert score, ETS and bias as they were reported by Keil and Craig (2009) for our calculations. $ASSIM_0$ is ASSIM before the correction for the location error, DAS is the *Displace and Amplitude Score* proposed by Keil and Craig (2009), ETS is the equitable threat score, and bias is the amplitude error. 19

3.3 Summary of error measures for the examples of Section 3.4.3. 21

5.1 Description of the 32 United States named winter storms. 36

1. INTRODUCTION

Increasing model resolution has led to qualitatively more realistic precipitation forecasts (Done et al., 2004; Romero et al., 2001; Speer and Leslie, 2002; Weisman et al., 1997). The greater qualitative realism of the forecast precipitation fields, however, does not necessarily translates into quantitatively more accurate forecasts. In fact, studies using traditional precipitation verification metrics have failed to show consistent improvements with increasing model resolution (e.g. Done et al., 2004; Mass et al., 2002). This result may be due to the limitations of the traditional, point-to-point verification techniques rather than to the lack of forecast improvements. In particular, such techniques indicate a large error in a situation where a generally well predicted precipitation event of high spatial variability is slightly misplaced.

A precipitation event of a specified time interval can be characterized by its *location*, the *total amount* of precipitation it produces, and the *structure* of the precipitation field (Wernli et al., 2008). Murphy (1995) argued that a verification approach designed for precipitation was conceptually and methodologically sound only if each identifiable aspect of forecast quality be measured seperately, and then absolute or relative performance be judged in the multidimensional space defined by these aspects of quality. The unique characteristics of precipitation and the difficulties to indepedently measure these characteristics of precipitation forecast have motivated us to design a verification technique that is able to automatically detect, quantify, and seperate location error from the other two error components for each precipitation event. The decompsed location, amplitude, and structure error components, once successfully measured, can be used to explore the typical limits of predictability of these three characteristics of precipitation events.

In a pair of papers, Keil and Craig (2007, 2009) (hereafter KC07 and KC09, respec-

tively) introduced a morphing technique for the verification of precipitation forecasts. Han and Szunyogh (2016) (hereafter HS16) proposed some algorithmic changes to the morphing technique of KC07 and KC09 and defined the morphing based measure of the location (position) error differently. In addition, while KC07 and KC09 described the other forms of error by a single measure, called the residual error, HS16 defined independent measures of the amplitude and the structure error.

This dissertation describes an improved morphing based technique for the verification of precipitation forecasts. The new technique provides accurate estimates of the components of the precipitation forecast error for a significantly longer range of the location error than the technique of HS16. It also introduces a new measure of the similarity between a pair of precipitation fields, called the *Amplitude and Structural Similarity Index Measure (ASSIM)*, which used to both define the convergence criterion for the revised (iterative) morphing technique and measure the amplitude and structure error. ASSIM is motivated by a series of papers (Wang and Bovik, 2002, 2009; Wang et al., 2004) that introduced a novel and widely popular *measure of image quality*. ASSIM is formally identical to the *Universal Quality Index (UQI)* of Wang and Bovik (2002) except for a minor modification that makes ASSIM positive semi-definite. The index is called ASSIM, in part, to acknowledge its formal similarity to the most often used form of the original image quality index, which is called *Structural Similarity Index Measure* (*SSIM*, Wang and Bovik, 2009; Wang et al., 2004), and in part, to emphasize that it also accounts for what is called amplitude error in the literature on precipitation forecast verification. SSIM was first proposed to be used as a similarity index of meteorological fields by Mo et al. (2014). The present study describes the first application of the measure to the verification of precipitation forecasts.

The improved morphing based technique is applied to ensemble precipitation forecasts. Specifically, the morphing based technique is used to explore the typical limits of predictability of the three characteristics (location, amplitude, structure) of precipitation

events for the named winter storms of the United States from the 2014/2015 and 2015/2016 storm seasons. The operational forecasts used for the study are global ensemble forecasts from the National Centers for Environmental Prediction (NCEP), while the verifying data are composed of Stage IV precipitation analyses and short-term forecasts from the European Centre for Medium Range Forecasts (ECMWF). The results provide information about, not only the predictability of the precipitation systems of winter storms, but also the performance of the NCEP ensemble forecasts in predicting such systems.

The structure of the dissertation is as follows. Chapter 2 is a literature review of the problems investigated in this dissertation. Chapter 3 summarizes the morphing based technique of HS16, describes the proposed changes to the technique, introduced ASSIM, and illustrates the usage of the proposed verification technique with idealized and realistic test examples. Chapter 4 introduces the methodology for the application of the proposed morphing based technique to ensemble precipitation forecasts. Chapter 5 describes the winter storm examples and the corresponding ensemble forecast and verification datasets. Chapter 6 presents and discusses the verification results. Chapter 7 offers the conclusions.

2. PRECIPITATION VERIFICATION METHODS

2.1 Spatial Verification Methods

2.1.1 Introduction

The goal of forecast verification is the characterization of the quality of the forecasts of the atmospheric fields (Wilks, 2011). The accuracy of a precipitation forecast is usually assessed by computing a measure of the correspondence between the gridpoint values $P^f_{i,j}$ and $P^a_{i,j}$, $i = 1, ..., M, j = 1, ..., N$ of the forecast precipitation field P^f and the corresponding verification field P^a. A precipitation forecast is perfect if $P^f_{i,j} = P^a_{i,j}$ for each gridpoint, $i = 1, ..., M, j = 1, ..., N$, of the verification domain. Traditional scalar summary scores, such as *mean-square-error (MSE)*, *root-mean-suqare error (RMS)*, and *threat score* have long been popular, because they are computationally straightforward. However, they provide incomplete information about the quality of a forecast because they only make comparison on a point-to-point basis with no regard to spatial information (Gilleland et al., 2009). Therefore, the results from traditional point-to-point verification methods are often difficult to interpret in meaningful physical terms.

Several new spatial verification methods have been developed over the last decade. These approaches treat each identifiable aspect of forecast quality seperately and judge the relative performance of precipitation forecasts in the multidimensional space defined by these aspects of quality. The one aspect of precipitation forecast quality that has gained a lot of attention is the location of a precipitation feature. Some techniques call the *location error* of a forecast precipitation features the *dispalcement error*. After the displacement error is seperated, intensity error and structure error can be defined.

The study of Hoffman et al. (1995) introduced the first verification approoach based on the concept of decomposing error in the forecasts of spatial fields into displacement,

amplitude and residual errors. The displacement and amplitude error were determined by applying a transformation function to the whole forecast field to optimize certain fit criterion of the forecast to the verifying data. The resulting transformed field was used to determine the residual error. This study inspired a new direction for the development of spatial verification techniques that are based on the spatial manipulation of the forecast field to resemble the verifying field as close as possible. The group of these techniques is often called *field deformation verification* or *morphing based verificiation*.

2.1.2 Object-based Verification

Several techniques have been proposed in the last decade to separate the displacement error from the other sources of errors, which are usually collectively referred to as *intensity error* (e.g., Davis et al., 2006; Ebert, 2009; Ebert and McBride, 2000; Gilleland et al., 2009). One group of approaches for the separation of the three error components is based on identifying distinct spatial patterns of precipitation as *objects*, and comparing the properties of the matching objects in the forecasts and the analyses (observations). The difference between the spatial locations of the matched objects is the *location error*, the difference between the spatial structures of the matched objects is the *structure error*, and the difference between the total amounts of precipitation in the matched objects is the *amplitude error*.

2.1.3 Morphing-based Verification

The technique proposed by Keil and Craig (2007, 2009) is a morphing based verification technique. The technique employs an *optical flow method* to morph the image (field) of the forecast precipitation into an image that resembles the image (field) of the analyzed precipitation as closely as possible. This method treats the precipitation as a passive scalar and carries out the morphing by computing a vector field, called the optical flow, that is then used to advect the original precipitation field. Because a successful morphing cor-

rects the displacement error, a measure of the displacement error can be defined based on the optical flow, and a measure of the intensity error can be defined by a measure of the difference between the two images.

2.1.4 Applications

Brewster (2003) extended the concept of displacement error to data assimilation applications. He proposed a technique to estimate and correct location errors (which he called it *phase correction*) of the background states estimate by using an algorithm that maximazes the correlation between the analyzed and observed features of the spatial scales of the storms. Brewster (2003) showed that the location error (phase) correction method was effective in producing an analysis field that agrees with the data, and preserved the structure developed by the model.

2.2 Ensemble Verification

2.2.1 Introduction

The practical solution to the analytic intractability of sufficiently detailed stochastic dynamic equations is to approximate these equations by using *Monte-Carlo methods*, as proposed by Epstein (1969) and Leith (1974), an approach now called *ensemble forecasting*. An ensemble forecast starts by drawing a finite sample from the probability distribution that describes the uncertainty of the initial state of the atmosphere. If the forecast model contains an accurate representation of the atmospheric dynamics, the dispersion of the ensemble obtained by integrating the model forward in time represents a random sample from the probability distribution function of the forecast uncertainty. For a properly constructed ensemble, the true state of the atmosphere would be just one more statistically indistinguishable member of the forecast ensemble (Anderson, 1997). Ensemble verification methods typically verify necessary conditions for this criterion.

2.2.2 Rank Histogram

The *rank histogram* (also called the *Talagrand Diagram*) is used to determine the extent to which the ensemble dispersion matches the dispersion of the distribution of the verifying observation (Anderson, 1996; Hamill and Colucci, 1997; Talagrand et al., 1997). Consider a K-member ensemble precipitation forecast of a fixed location and a corresponding observed value of the same location. The rank of the observed value is determined by comparing it to the corresponding K forecast values. For example, if the observed value is smaller than any of the K values for the ensemble, its rank is $i = 1$. For a collection of n such forecast ensembles and the corresponding n independent, single observed values, the probability of the $K+1$ different values of the rank of the observed value can be estimated and plotted. The constructed histogram is called the *rank histogram*. If the K members and the single observation are drawn from the same distribution, the rank of the observed value is equally likely to be $i = 1, 2, .., K$, or $K + 1$. Therefore, a necessary condition that the ensemble has to satisfy in order to be considered consistent with the observations is that the histogram of the verification rank is uniform, for a large sample. Departures from rank uniformity can be used to diagnose aggregate deficiencies of the ensembles (Hamill, 2001). The interpretation of the rank histogram is straightfoward if the verifying variable is nearly normally distributed. However, for variables like precipitation, whose distribution is highly skewed, the interpretation of the rank histogram is more likely to be misleading (Casati et al., 2008).

2.2.3 Continuous Rank Probability Score (CRPS)

Another widely used metrics for the verification of ensemble forecasts is the *continuous rank probability score* (CRPS). CRPS measures the difference between the cumulative distribution function (CDF) of the ensemble and the observation of a scalar variable. The CDF of the observation, a point value, is a Heaviside (step) function with the step at the

value of the observation. The CDF of the forecast is a piecewise constant function defined according the values of the predicted variable of each ensemble member (Candille et al., 2007; Casati et al., 2008; Stanski et al., 1989; Wilks, 2011). CRPS summarizes the quality of the ensemble by integrating the squared difference between the forecast CDF and the observed CDF over all possible thresholds of the variable under consideration.

3. THE PROPOSED VERIFICATION TECHNIQUE

Assume that the precipitation fields are represented by a collection of d-by-d elementary pixels: each pixel represents the areal mean of the total precipitation for the pixel for a specific time interval. Given a forecast precipitation field P^f and a corresponding verifying analysis field P^a, the goal is to morph the image of P^f into an approximate image of P^a. The morphing is carried out by computing a morphing vector for each (elementary) pixel of P^f to move the pixels to their new locations in the morphed image. The location error is defined by the mean of the morphing vectors, and the other error components are computed after a correction of P^f for the location error.

3.1 The Original Morphing Technique

The morphing techniques of KC07, KC09, and HS16 all compute the morphing vectors by a pyramid matching algorithm, in which the number of pyramid levels F is called the sub-sampling parameter. At level k ($k = F, F-1, \ldots, 0$) of the algorithm, the forecast and the verifying precipitation fields are coarse-grained by averaging $2^k \times 2^k$ (elementary) pixels of both fields. Then, each pixel of the coarse-grained forecast image is shifted by $0, 1, -1$ position in both the zonal (x) and meridional (y) direction. The dx zonal and dy meridional component of the morphing vector $d\mathbf{X}$ is chosen to be the pair of values of $0, 1, -1$ that minimizes the absolute value of the difference between the coarse-grained forecast pixel and the coarse-grained verifying analysis pixel. (The same $d\mathbf{X}$ is assigned to all $2^k \times 2^k$ elementary pixels that make up a coarse grained pixel, but $d\mathbf{X}$ can be different for the different coarse-grained pixels.) The procedure is repeated until its completion for the last level (level 0), at which the coarse grained pixels are identical to the elementary pixels ($2^0 \times 2^0 = 1$).

There are a number of small but important differences in the technical details of the

implementation of the pyramid matching algorithm by KC07, KC09, and HS16. Most importantly, HS16 imposes a constraint at the top level of the pyramid to prevent moving pixels of P^f in which the precipitation is not likely to be related to a precipitation feature of P^a (see Section 2.c of HS16 for details). This constraint is also imposed in the calculations of Section 3.4 of the dissertation.

The maximum total distance that an elementary pixel of the forecast field can move as the result of morphing in either direction is $2^F \times d$. Unlike the techniques of KC07 and KC09, the technique of HS16 does not require the specification of F, as it chooses the optimal value of F based on P^f and P^a (see Section 3.b of HS16 for details). The technique requires, however, the specification of the maximum allowable value F_{max} of F. The main motivations to cap the value of F are to limit the distance $L_{max} = 2^{F_{max}} \times d$ in both the zonal and meridional directions within which precipitation features can be considered matching and reduce the likelihood of the presence of multiple precipitation systems within the search distance. Limiting the value of F, however, has an unintended consequence: it leads to an underestimation of the location error in situations where only part of the forecast feature is within distance L_{max} from the verifying feature. This dissertation proposes a number of simple modifications to the morphing technique to correct this flaw. The new technique is also computationally more efficient than the original one, which is an important adva ntage when a large number of cases have to be processed.

3.2 The Modified Morphing Technique

In the proposed version of the morphing technique, the pyramid matching algorithm is applied iteratively. While it still has the parameter F_{max}, it can move pixels to a maximum distance that is significantly longer than L_{max}.

The steps of the algorithm are the following:

1. *(Filtering)* Prepare a filtered image $P^f_{filtered}$ of P^f by replacing each non-zero pixel

of P^f that is at least L_{max} pixels apart from the nearest non-zero pixel of P^a in either direction by a zero pixel.

2. *(Morphing)* Use the pyramid matching algorithm to morph $P^f_{filtered}$ into P^a to obtain the field of morphing vectors $d\mathbf{X}_{i,j}$, where $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$ identify the location of a pixel of $P^f_{filtered}$ in the x and y direction, respectively.

3. *(Shifting)* Compute the mean

$$d\mathbf{X}_{mean} = \frac{1}{n} \sum_{i=1}^{M} \sum_{j=1}^{N} d\mathbf{X}_{i,j} \tag{3.1}$$

of the morphing vectors, where n is the number of nonzero pixels of $P^f_{filtered}$. Round the dx_{mean} and dy_{mean} components of $d\mathbf{X}_{mean}$ to the nearest integers and shift all nonzero pixels of P^f by $d\mathbf{X}_{mean}$.

4. *(Iteration)* Replace P^f by the shifted field and repeat Steps 1, 2, and 3 unitl the shifted field matches P^a as closely as possible (a formal criterion for "matching P^a as closely as possible" will be provided in Section 3.3).

The key modification of the morphing process is the repeated (iterative) application of the pyramid matching algorithm and the shift vector $d\mathbf{X}_{mean}$. (In HS16, the forecast feature was shifted by $d\mathbf{X}_{mean}$ only once, after the completion of the morphing process, as it was part of the computation of the structure error measure rather than the morphing technique.) The final shifted forecast field can be equivalently computed by applying the sum of $d\mathbf{X}_{mean}$ from all iterations to the pixels of the original forecast field P^f.

3.3 ASSIM

ASSIM is defined to both provide a formal definition of "matching P^a as closely as possible given P^f" in step 4 of the morphing technique and measure the amplitude and

structure error. ASSIM has three components: the *amplitude distortion* (D_{amp}), *variance distortion* (D_{var}), and *loss of correlation* (D_{cor}). The amplitude distortion is a measure of the error in the prediction of the total precipitation, the variance distortion is a measure of the error in the variability (spatial structure) of the precipitation field, while the loss of correlation is a measure of the loss of point-wise correspondence between the forecast and the analyzed precipitation amount.

Formally, the amplitude distortion is defined by

$$D_{amp}\left(P^f, P^a\right) = \frac{2\mu_f\mu_a}{\mu_f^2 + \mu_a^2}, \tag{3.2}$$

where $\mu_f = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{i,j}^f$ and $\mu_a = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{i,j}^a$ are the mean forecast precipitation and the mean analyzed precipitation, respectively. The variance distortion is defined by

$$D_{var}\left(P^f, P^a\right) = \frac{2\sigma_f\sigma_a}{\sigma_f^2 + \sigma_a^2}, \tag{3.3}$$

where σ_f is the standard deviation of the forecast precipitation over all pixels, and σ_a is the standard deviation of the analyzed precipitation. The loss of correlation is measured by the correlation between P^f and P^a,

$$D_{cor}\left(P^f, P^a\right) = \begin{cases} \frac{\sigma_{fa}}{\sigma_f\sigma_a}, & \text{if } \sigma_{fa} \geq 0, \\ 0, & \text{if } \sigma_{fa} < 0. \end{cases} \tag{3.4}$$

where σ_{fa} is the covariance

$$\sigma_{fa} = \frac{1}{MN-1}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(P_{i,j}^f - \mu_f\right)\left(P_{i,j}^a - \mu_a\right) \tag{3.5}$$

between P^f and P^a. ASSIM is defined by

$$D\left(P^f, P^a\right) = \left[D_{amp}\left(P^f, P^a\right)\right]^{\alpha} \cdot \left[D_{var}\left(P^f, P^a\right)\right]^{\beta} \cdot \left[D_{cor}\left(P^f, P^a\right)\right]^{\gamma}, \qquad (3.6)$$

where α, β, and γ ($\alpha + \beta + \gamma = 1$) are non-negative real numbers specified by the user.

$D(P^f, P^a)$ satisfies the following conditions:

- *symmetry*: $D(P^f, P^a) = D(P^a, P^f)$.
- *boundedness*: $0 \leq D(P^f, P^a) \leq 1$.
- *unique maximum*: $D(P^f, P^a) = 1$, if and only if $P^f_{i,j} = P^a_{i,j}$ for $i = 1, \ldots, M, j = 1, \ldots, N$.

The definition of ASSIM given by Eq. (3.6) is formally identical to that of UQI (Wang and Bovik, 2002), except for the definition of $D_{cor}\left(P^f, P^a\right)$. In the definition of UQI, this term is simply defined by the correlation between the two fields; thus, unlike $D_{cor}\left(P^f, P^a\right)$, which can take a value in the closed interval $[0, 1]$, it can take a value in the closed interval $[-1, 1]$. Wang et al. (2004) later modified the definition of UQI to include a small additive constant in both the numerators and denominators of the three terms whose product defined UQI and called the resulting measure SSIM. The goal with adding the small constants was to eliminate the singularity that arose when the denominators of the ratios were very small. Such singularity does not occur for the precipitation forecasts studied in this dissertation. Should such singularity arise in future applications, the value of $D_{amp}\left(P^f, P^a\right)$ and $D_{amp}\left(P^f, P^a\right)$ could be set to zero and the value of $D_{var}\left(P^f, P^a\right)$ to one in cases of singularity.

For a precipitation forecast that provides a perfect prediction of the location, amplitude and structure, ASSIM takes the value of 1. A non-zero location error reduces ASSIM, because it reduces $D_{cor}\left(P^f, P^a\right)$. Thus a morphed forecast field "matches P^a as closely as

possible given P^f" when ASSIM takes its maximum for P^f and P^a. In practical terms, the iterations of the morphing technique can be stopped once an iteration fails to increase ASSIM by more than a prescribed small value ε. In addition, if the estimate of the maximum of ASSIM obtained by the morphing technique is smaller than a predefined threshold δ, we can declare that there is no precipitation feature in the forecast that would match the precipitation feature in the analysis.

3.4 Illustration of the Verification Technique

This section illustrates the behavior of the proposed verification technique by applying it to an idealized example of HS16, idealized and realistic examples of the Spatial Verification Methods Intercomparison Project (ICP) (Ahijevych et al., 2009), and winter storm forecasts from National Centers for Environmental Prediction (NCEP). In the calculations presented here the three properties of the amplitude and structural errors are assumed to be equally important for the hypothetical forecast user, choosing the parameter values to be $\alpha = \beta = \gamma = 1/3$.

3.4.1 Idealized Example

In this example, a series of idealized forecasts with increasing location error. The precipitation feature is an 8-by-8 pixel square, which is initially located in the bottom left corner of the verification region in both P^f and P^a (Fig. 3.1). The precipitation feature of P^f is then shifted gradually (one pixel at a time) in both the x and y direction to create the series of simulated forecasts. The panels of Fig. 3.1 illustrate the effect of each step of the first two iterations of the morphing technique for the forecast in which the precipitation feature is misplaced by 12 pixels in both the x and y direction. In this calculation, $F_{max} = 3$ and the iteration is stopped once ASSIM becomes 1. The first filtering step (top left panel) filters the 3/4 part of the forecast precipitation feature that is outside of the search region, that is, beyond a $L_{max} = 2^{F_{max}} = 8$ pixel distance from

the border of the analysis feature in either the zonal or the meridional direction. The first application of the pyramid matching algorithm (top middle) moves the remaining 1/4 of the forecast precipitation feature toward the verifying precipitation feature, while the he first shift step (top right) moves the entire forecast precipitation feature into the search region. The steps of the second iteration (bottom panels) result in a perfect match of the two precipitation features.

Figure 3.2 illustrates the change in the behavior of the morphing technique due to the proposed modifications. It shows that when the forecast feature is misplaced by $2^{F_{max}} = 8$ or fewer pixels, both techniques capture the location error correctly. But, while the modified technique provides a correct estimate of the location error up to a displacement of 15 pixels, the original technique underestimate the location error at a displacement of 9 or more pixels.

3.4.2 ICP Cases

3.4.2.1 Idealized Cases

The proposed verification scheme is applied to 5 idealized ICP examples. In these examples, whose graphical illustration can be found in Ahijevych et al. (2009), both the forecast and analysis features have elliptical shapes. The forecast errors are errors in the location, area, and aspect ratio of the ellipsoids. Table 3.1 summerizes the estimates of $d\mathbf{X}$ and ASSIM for these examples. The results show that when the forecasts only have location error (cases 1 and 2), the verification technique does a perfect job: the estimate of the location error has no error and both ASSIM and its components correctly indicate that there is no amplitude or structure error. When both location, and amplitude and structure error are present, the estimates of the location error can be slightly less accurate. In particular, while the technique correctly detects that there is no location error in the y direction for cases 3 and 4, it estimates the location errors in the x direction with a relative error of

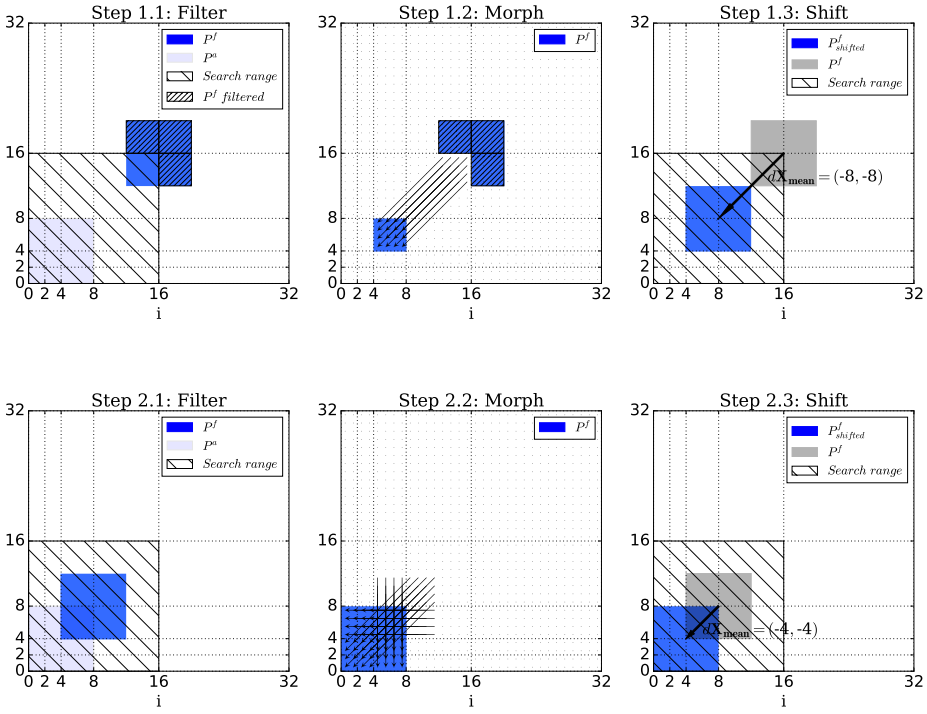

Figure 3.1: Illustration of the steps of the proposed morphing technique for an idealized case in which the forecast has only location error (see Section 3.4.1 for details).

Table 3.1: Summary of the results for the idealized ICP cases.

Case	Description of error	$d\mathbf{X}[(pixel, pixel)]$	D_{amp}	D_{var}	D_{cor}	D
1	50-pixel displacement in x direction	(-50,0)	1	1	1	1
2	200-pixel displacement in x direction	(-200,0)	1	1	1	1
3	125-pixel displacement in x direction, area error	(-129,0)	0.468	0.812	0.491	0.571
4	125-pixel displacement in x direction, wrong aspect ratio	(-121,0)	1	1	0.31	0.677
5	125-pixel displacement in x direction, large area error	(-125,0)	0.245	0.661	0.301	0.365

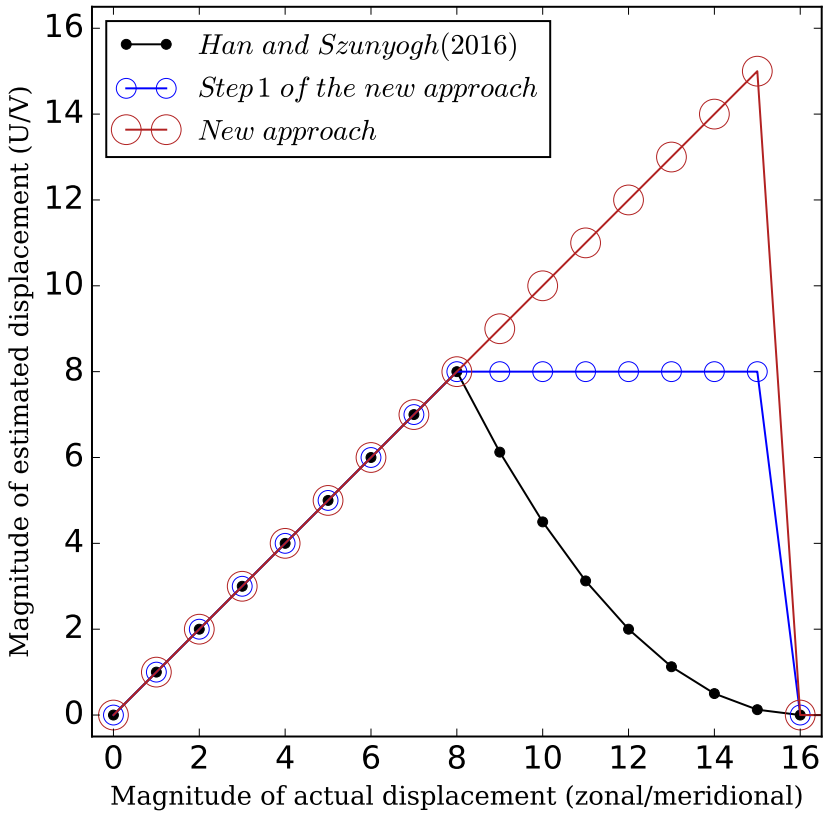

Figure 3.2: The estimates of the displacement errors for the series of simulated forecasts described in Section 3.4.1 that were obtained by (black) the technique of HS16 and (red) the proposed technique of the present paper.

3.2%. When there is no error in the amplitude and the only source of structure error is an error in the aspect ratio (case 4), the components of ASSIM correctly indicate that there is no amplitude error and the variance of the precipitation field is accurately predicted.

3.4.2.2 Realistic Cases

The proposed verification scheme is applied to all 27 forecasts of ICP and the results are summarizes in Table 3.2. Before turning the attention to the description of the results of the table, the behavior of the technique is illustrated on one of the 27 forecasts (Fig.3.3). As for all other forecasts in the data set, P^f is a 4-km resolution, 24-hour forecast of the 1-hour accumulation and P^a is a stage II precipitation analysis (Lin and Mitchell, 2005). Both the forecast and the analysis (left and middle panel) display a north-south oriented precipitations system located in the center of the verification domain along with scattered rain to the southeast. The main rainband, however, is slightly misplaced to the west in the forecast. The shifted forecast produced by the verification technique (right panel) almost completely removes the location error.

A useful feature of the ICP data set is that expert scores are available for all 27 forecasts (Ahijevych et al., 2009). These expert scores are the averages of scores assigned by each member of a panel of 26 experts. The numbers in Table 3.2 are the correlations between the expert scores and some selected objective scores including ASSIM. The set of 27 forecasts of ICP consists of forecasts from three models for nine forecast cases. The second column shows the correlations for all 27 forecasts, while the third column shows the correlations obtained by first computing the correlations for each model and then averaging the results over the different models. The numbers indicate that the correlation is almost always higher for ASSIM than the other objective scores. (The only exception is the higher correlation for DAS for the NCEP model.)

A comparison of the first two rows of the table confirms that correcting for the location

Table 3.2: The correlation between the different objective measures and the subjective expert score for the realistic ICP forecasts. We use the values of the expert score, ETS and bias as they were reported by Keil and Craig (2009) for our calculations. ASSIM_0 is ASSIM before the correction for the location error, DAS is the *Displace and Amplitude Score* proposed by Keil and Craig (2009), ETS is the equitable threat score, and bias is the amplitude error.

Correlation for	all 27 data points	9 cases averaged over model	CAPS WRF model	NCAR WRF model	NCEP WRF model
ASSIM_0	0.526	0.498	0.529	0.636	0.329
ASSIM	0.614	0.598	0.650	0.711	0.434
DAS	0.255	0.146	0.368	0.028	0.778
ETS	0.410	0.346	0.491	0.523	0.023
Bias	0.467	0.439	0.489	0.432	0.395

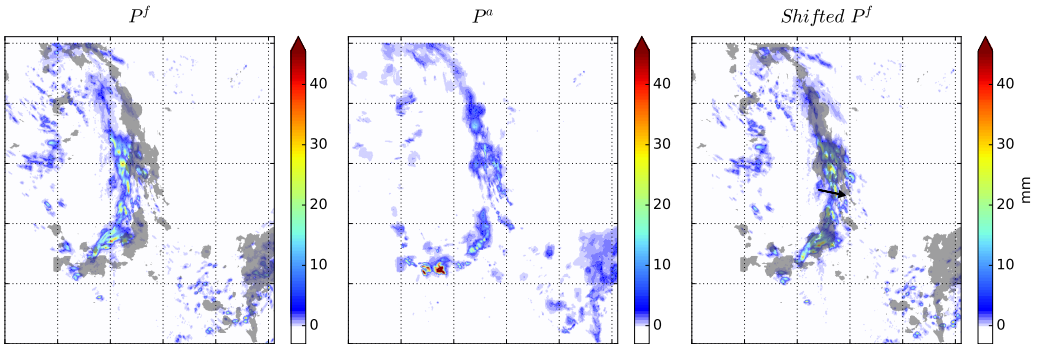

Figure 3.3: Illustration of a realistic ICP example. Shown are (left) an 24 h forecast of the 1 h precipitation total by the NCEP WRF model, (middle) the related Stage II analysis of 1 h accumulation, and (right) the shifted forecast obtained by the proposed morphing technique. The black arrow in the right panel indicates the magnitude and direction of $d\mathbf{X}_{mean}$ (the total shift vector).

error improves the correspondence between ASSIM and the subjective expert scores. In addition, a comparison of the second row to the rows that follow reveals that with the exception of one case (NCEP WRF and DAS), the correlation is higher for the ASSIM corrected for location error than for the other verification scores. This result supports the argument of Wang and Bovik (2009) that a measure of this type better reflects the property of the human visual system that it is less sensitive to small displacement errors than the other objective quality scores. It should be emphasized, however, that in the case of precipitation systems, this advantageous property of the measure can be realized only after a correction for the location error.

3.4.3 Winter Storms

Finally, the performance of the verification technique is tested on operational precipitation forecasts of a couple of United States winter storms from the 2014/2015 and 2015/2016 winter storm season. The operational forecasts are global medium range forecasts from the National Centers for Environmental Prediction (NCEP), while the verifying analyses are operational 6-h precipitation forecasts from the European Centre for Medium Range Forecasts (ECMWF).

The left and middle panel of Fig. 3.4 shows a precipitation forecast with large location error. The proposed recursive morphing algorithm removes the displacement error by shifting the forecast feature in three iterations (right panel). The value of ASSIM before the correction of the displacement error is $D\left(P^f, P^a\right) = 0$ (Table 3.3) and reaches its maximum value of 0.851 after the third iteration of the morphing-based technique. This result indicates that despite the large location error, this 240 h forecast provides a surprisingly high quality prediction of the amplitude and structure of the precipitation event. This skill of the forecast could not have been detected without the prior correction of the location error.

Table 3.3: Summary of error measures for the examples of Section 3.4.3.

Storm	$\|d\mathbf{X}\|$ [km]	ASSIM before the correction of the location error	ASSIM after the correction of the location error
240h forecast of Winter Storm Cato	1140	0	0.851
120h forecast of Winter Storm Sparta	490	0.553	0.842
360h forecast of Winter Storm Sparta	-	0	0.657

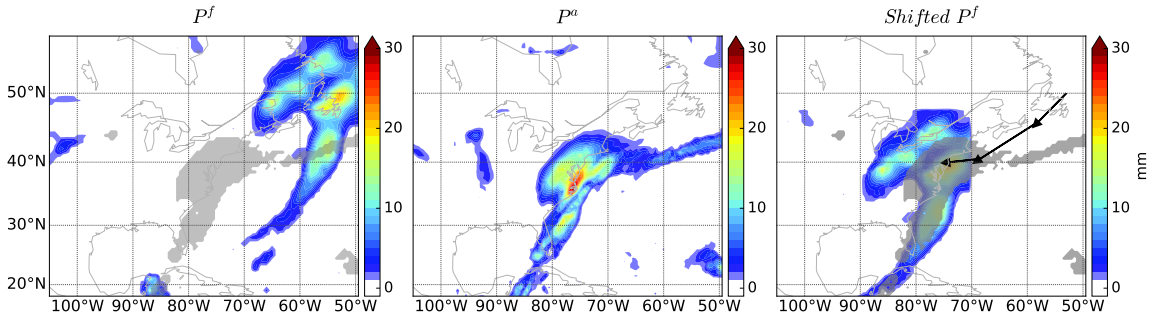

Figure 3.4: Illustration of the example of Winter Storm Cato. (left) 240 h NCEP forecast of the 6 h precipitation total at 0000 UTC November 26, 2014, (middle) the verifying analysis, and (right) the forecast shifted by $d\mathbf{X}_{mean}$. The grey shading in the left and right panel indicates the contour of the precipitation field in the verifying analysis (middle panel). The black arrows in the right panel are the shift vectors of the iterations of the morphing algorithm.

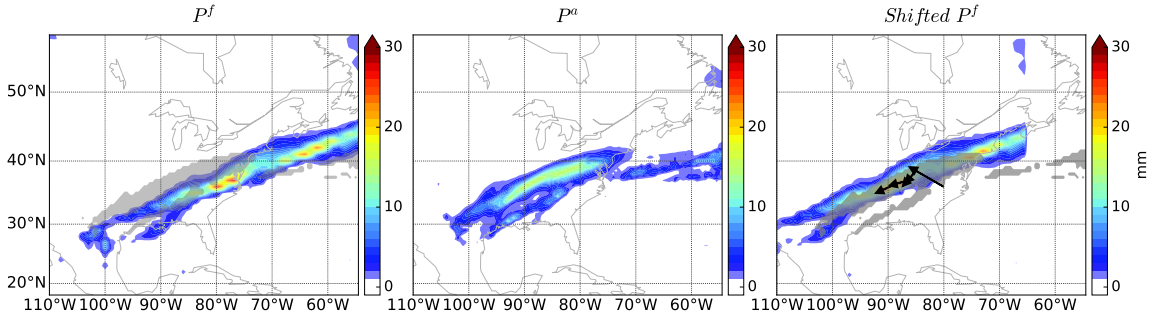

Figure 3.5: Same as Fig 3.4, except for the 120 h forecast of winter storm Sparta at 0000 UTC, March 05, 2015.

Figure 3.5 shows another example to illustrate the behavior of verification technique. A subjective visual inspection of the figure could reasonably lead to the conclusion that the forecast is of high quality. However, the value of ASSIM before the correction of the location error is small (0.553). After the correction of the location error, in five iterations, we obtain $D\left(P^f, P^a\right) = 0.842$. The smaller value of ASSIM for the forecast of Fig. 3.5 than for that of Fig. 3.4 indicates that the amplitude and structure error of the 5-day forecast of Fig. 3.5 is larger than that of the 10-day forecast of Fig. 3.4.

Figure 3.6 is an example of a precipitation forecast for which the proposed technique indicates that it failed to predict the verifying event. In this example, the morphing technique shifts the north-south oriented forecast precipitation system in an attempt to match the northeast-southwest oriented rainband of the verifying system. The value of ASSIM increases from 0 to 0.657 after the shift, but does not reach the threshold value $\delta = 0.8$ that we use to declare a forecast feature a match.

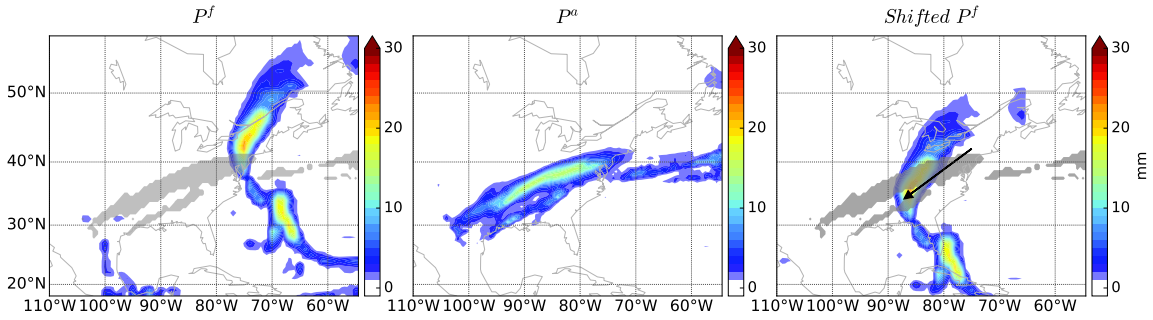

Figure 3.6: Same as Fig 3.4, except for the 360 h forecast of winter storm Sparta at 0000 UTC, March 05, 2015.

4. METRICS OF UNCERTAINTY FOR THE VERIFICATION OF ENSEMBLE PRECIPITATION FORECASTS

Given is a K-member ensemble of precipitation forecasts for each winter storm. A precipitation analysis that is assumed to represent the true precipitation field at the resolution of the forecasts is also available for each storm. The proposed morphing based technique is applied to each member of each ensemble forecast to determine whether or not the particular ensemble member predictes the storm. If the member predicts the storm, the algorithm also provides a vector $d\mathbf{X}$ that represents the spatial shift required to correct for the location error. The reliability of the ensemble in predicting the uncertainty in the location of the precipitation events is assessed based on the ensembles of $d\mathbf{X}$. The reliability of the ensemble in predicting the uncertainty of the amplitude and structure of the precipitation system of the storm is also assessed.

4.1 Finding a Matching Forecast System

The iterative morphing-based algorithm of Chapter 3 seeks the shift vector $d\mathbf{X}$ that maximizes the similarity between the shifted forecast precipitation system and the precipitation system of the verifying event. The similarity between the precipitation field of the shifted forecast, $P^f_{shifted}$, and the precipitation field of the verifying analysis, P^a is measured by ASSIM.

As explained in Chapter 3, ASSIM has two functions in the proposed morphing based verification strategy. First, the iterations of the morphing technique are stopped once an iteration fails to increase ASSIM by more than a prescribed small value ε. Second, if the estimate of the maximum of ASSIM obtained by the morphing technique is smaller than a predefined threshold δ, it is declared that there is no precipitation feature in the forecast that would match the precipitation feature in the analysis. In this chapter, ASSIM

is used to define the minimum level of similarity that is acceptable to declare that a forecast precipitation system is a prediction of the precipitation system of a particular winter storm.

4.2 Position Forecasts

4.2.1 Definition of the Position Uncertainty

Let K' be the number of ensemble members that predict a particular storm, that is, the number of ensemble members for which $D\left(P^f_{shifted}, P^a\right)$ is larger than a prescribed value D_0. The value of K' depends on the storm and the forecast lead time, but it always satisfies the relationship $K' \leq K$ for a K-member ensemble. For a particular storm, the value of K' typically decreases with the forecast time.

Denote the value of $d\mathbf{X}$ for the k-th ensemble member that predicts the storm by $d\mathbf{X}^k$ ($k = 1, 2, \ldots, K'$). One can think of $d\mathbf{X}^k$ as the difference between the location $\mathbf{r}^k_f$ of the precipitation feature in the k-th ensemble member and the location $\mathbf{r}_a$ of the same precipitation feature in the verifying analysis. The vectors of positions $\mathbf{r}^k_f$, $k = 1, 2, \ldots, K'$, and $\mathbf{r}_a$ are introduced only to develop ideas, as the resulting metrics will not assume their explicit knowledge.

Formally,

$$\mathbf{r}^k_f + d\mathbf{X}^k = \mathbf{r}_a, \tag{4.1}$$

where $\mathbf{r}^k_f$ could be defined by the position of any point of the forecast feature in the k-th ensemble member. Taking the ensemble mean of Eq. (4.1) and rearranging the result yields

$$\overline{d\mathbf{X}} = \mathbf{r}_a - \bar{\mathbf{r}}_f, \tag{4.2}$$

where

$$\begin{aligned} \overline{d\mathbf{X}} &= \frac{1}{K'}\sum_{k=1}^{K'} d\mathbf{X}^k, \\ \bar{\mathbf{r}}_f &= \frac{1}{K'}\sum_{k=1}^{K'} \mathbf{r}_f^k. \end{aligned} \tag{4.3}$$

Equation (4.2) shows that $\overline{d\mathbf{X}}$ is equal to the error in the ensemble mean prediction $\bar{\mathbf{r}}_f$ of the position $\mathbf{r}_a$.

Ensemble forecasting and verification techniques treat the forecast parameters as random variables, with the sources of randomness being the random errors of the initial conditions and the models. Following this approach, the position $\mathbf{r}$ is also considered a random variable, $\mathbf{r}_a$ is a realization of $\mathbf{r}$, and $\bar{\mathbf{r}}_f$ is a prediction of the mean $\bar{\mathbf{r}}$ of $\mathbf{r}$. According to Eq. (4.2), $\overline{d\mathbf{X}}$ is the difference between a realization $\mathbf{r}_a$ of a random variable and a prediction $\bar{\mathbf{r}}_f$ of the mean of that random variable. In addition, Eq. (4.2) can be written as

$$\overline{d\mathbf{X}} = \boldsymbol{\epsilon}_{pos} - \mathbf{b}_{pos}, \tag{4.4}$$

where

$$\mathbf{b}_{pos} = \bar{\mathbf{r}}_f - \bar{\mathbf{r}} \tag{4.5}$$

is the error in the prediction of the mean,

$$\bar{\mathbf{r}} = E\left[\mathbf{r}\right] \tag{4.6}$$

and

$$\boldsymbol{\epsilon}_{pos} = \mathbf{r}_a - \bar{\mathbf{r}} \tag{4.7}$$

is a realization of the random variable $\mathbf{r} - \bar{\mathbf{r}}$, which is a representation of the *analysis*

uncertainty. Here, $E[\cdot]$ is the expected value of the probability distribution of $\mathbf{r}$ given all sources of forecast uncertainty.

Ideally, the probability distribution sampled by the ensemble should be equal to the true probability distribution. Ensemble verification techniques verify conditions that are necessary for the ensemble to satisfy to have this desired property, because the forecast performance of an ensemble has to be typically assessed based a single verification data for each forecast ($\mathbf{r}_a$ in the present case). This can be done by verifying that the predicted error statistics are consistent with the actual error statistics for a sufficiently large sample of forecasts. For example, let $\bar{\mathbf{r}}_f^m$, $m = 1, 2, \ldots, M$, be a sample of the predicted mean positions and $\mathbf{r}_a^m$, $m = 1, 2, \ldots, M$, the corresponding verification data. If the ensemble members and the verification data are drawn from the same distribution, the expected value

$$E\left[\frac{1}{M}\sum_{m=1}^{M}\bar{\mathbf{r}}_f^m\right] \tag{4.8}$$

of the sample mean of the predicted means is equal to the expected value

$$E\left[\frac{1}{M}\sum_{m=1}^{M}\mathbf{r}_a^m\right] \tag{4.9}$$

of the sample mean of the verification data, because then

$$E\left[\frac{1}{M}\sum_{m=1}^{M}\mathbf{r}_a^m\right] = \frac{1}{M}\sum_{m=1}^{M}E\left[\mathbf{r}_a^m\right] = \frac{1}{M}\sum_{m=1}^{M}E\left[\mathbf{r}_f^{mk}\right] = \frac{1}{M}\sum_{m=1}^{M}E\left[\bar{\mathbf{r}}_f^m\right] = E\left[\frac{1}{M}\sum_{m=1}^{M}\bar{\mathbf{r}}_f^m\right]. \tag{4.10}$$

The second step made use of the assumption that $\mathbf{r}_a^m$ and $\mathbf{r}_f^{mk}$, where $\mathbf{r}_f^{mk}$ is the position of ensemble member k in forecast m, are drawn from the same probability distribution. The third step took advantage of the property of the ensemble mean $\bar{\mathbf{r}}_f^m$ that it is an unbiased estimate of the mean of the distribution from which $\mathbf{r}_f^{mk}$, $k = 1, 2, \cdots, K(m)'$ are drawn,

independently of the value of $K'(m)$. In addition, because for a sufficiently large M, the variance of the mean of the M samples is small,

$$\frac{1}{M}\sum_{m=1}^{M}\bar{\mathbf{r}}_f^m \approx \frac{1}{M}\sum_{m=1}^{M}\mathbf{r}_a^m. \tag{4.11}$$

Equation (4.11) is a necessary condition that an ensemble satisfies, if it samples the (true) probability distribution of the forecast uncertainty.

4.2.2 Position Bias

Because $\mathbf{r}_a$ is a realization of $\mathbf{r}$,

$$E[\boldsymbol{\epsilon}_{pos}] = E\left[\mathbf{r}_a - \bar{\mathbf{r}}\right] = \mathbf{0}. \tag{4.12}$$

Thus taking the expected value of Eq. (4.4) yields

$$E\left[\overline{d\mathbf{X}}\right] = E[\mathbf{b}_{pos}], \tag{4.13}$$

which by analogy with Eqs. (4.10) and Eqs. (4.11) leads to

$$\frac{1}{M}\sum_{m=1}^{M}\overline{d\mathbf{X}}^m \approx \frac{1}{M}\sum_{m=1}^{M}\mathbf{b}_{pos}^m. \tag{4.14}$$

for a sufficiently large M. The left-hand side of Eq. (4.14) is an estimate of the climatological mean of $\mathbf{b}_{pos}$, which we will refer to as the *position bias*.

4.2.3 Position Spread

If the ensemble samples the (true) probability distribution of $\mathbf{r}$ (e.g., Leith, 1974),

$$E\left[\frac{1}{K'-1}\sum_{k=1}^{K'}\left(\mathbf{r}_f^k - \bar{r}_f\right)^2\right] = E\left[\left(\mathbf{r}-\bar{\mathbf{r}}\right)^2\right]. \tag{4.15}$$

Next, it will be shown that both sides of Eq. (4.15) can be estimated based on the M-member sample of $d\mathbf{X}^k$, $k = 1, 2, \ldots, K'$, which we start by rewriting the right-hand side as

$$\begin{aligned} E\left[(\mathbf{r} - \bar{\mathbf{r}})^2\right] &= E\left[(\mathbf{r} - \bar{\mathbf{r}}_f - \bar{\mathbf{r}} + \bar{\mathbf{r}}_f)^2\right] \\ &= E\left[(\mathbf{r} - \bar{\mathbf{r}}_f)^2\right] + E\left[(\bar{\mathbf{r}}_f - \bar{\mathbf{r}})^2\right] \\ &= E\left[(\mathbf{r} - \bar{\mathbf{r}}_f)^2\right] + \frac{1}{K'} E\left[(\mathbf{r} - \bar{\mathbf{r}})^2\right]. \end{aligned} \tag{4.16}$$

The second step made use of the statistical independence of $\mathbf{r} - \bar{\mathbf{r}}_f$ and $\bar{\mathbf{r}}_f - \bar{\mathbf{r}}$, and in the third step, the well-known result of statistics that the variance of the estimation error $\bar{\mathbf{r}}_f - \bar{\mathbf{r}}$ that is due to sample fluctuations satisfies the equation

$$E\left[(\bar{\mathbf{r}}_f - \bar{\mathbf{r}})^2\right] = \frac{1}{K'} E\left[(\mathbf{r} - \bar{\mathbf{r}})^2\right]. \tag{4.17}$$

Rearranging Eq. (4.16) yields

$$E\left[(\mathbf{r} - \bar{\mathbf{r}})^2\right] = \frac{K'}{K' - 1} E\left[(\mathbf{r} - \bar{\mathbf{r}}_f)^2\right]. \tag{4.18}$$

It can be seen by making use of Eq. (4.3) that

$$\delta_{pos}^2 = \frac{1}{M} \sum_{m=1}^{M} \left(\Delta_{pos}^2\right)^m, \tag{4.19}$$

where

$$\left(\Delta_{pos}^2\right)^m = \frac{K'}{K' - 1} \left(\overline{d\mathbf{X}}^2\right)^m, \tag{4.20}$$

is an estimate of the right-hand side, and therefore also the left-hand side, of Eq. (4.18).

Now that we have a formula to estimate the right-hand side of Eq. (4.15), we turn our

attention to The next task is to find a formula for the estimation of the left-hand side of Eq. (4.15). An estimate analogous to Eq. (4.20) is

$$\sigma_{pos}^2 = \left[\frac{1}{M} \sum_{m=1}^{M} \left(\Sigma_{pos}^2 \right)^m \right], \tag{4.21}$$

where

$$\Sigma_{pos}^2 = \frac{1}{K' - 1} \sum_{k=1}^{K'} \left(\mathbf{r}_f^k - \bar{\mathbf{r}}_f \right)^2. \tag{4.22}$$

The *ensemble spread of the position* $(\Sigma_{pos})^m$ for sample m can be computed by

$$\left(\Sigma_{pos}^2 \right)^m = \frac{1}{K' - 1} \sum_{k=1}^{K'} \left(\overline{d\mathbf{X}}^m - d\mathbf{X}^{mk} \right)^2, \tag{4.23}$$

because

$$\begin{aligned} \Sigma_{pos}^2 &= \frac{1}{K' - 1} \sum_{k=1}^{K'} \left(\mathbf{r}_f^k - \bar{\mathbf{r}}_f \right)^2 = \frac{1}{K' - 1} \sum_{k=1}^{K'} \left(\mathbf{r}_a - d\mathbf{X}^k + \overline{d\mathbf{X}} - \mathbf{r}_a \right)^2 \\ &= \frac{1}{K' - 1} \sum_{k=1}^{K'} \left(\overline{d\mathbf{X}} - d\mathbf{X}^k \right)^2. \end{aligned} \tag{4.24}$$

In summary, the condition defined by Eq. (4.15) can be verified by checking that

$$\sigma_{pos} \approx \delta_{pos}. \tag{4.25}$$

4.3 Amplitude Forecasts

Let μ_f^k be the areal mean of the forecast precipitation in the verification domain for the k-th ensemble member, $k = 1, ..., K'$, and μ_a the areal mean of precipitation in the verification domain for the verifying analysis. The ensemble mean of the areal mean of

the forecast precipitation is

$$\bar{\mu}_f = \frac{1}{K'} \sum_{k=1}^{K'} \mu_f^k. \tag{4.26}$$

The analogue of Eq. (4.4) for the areal mean of the precipitation is

$$\mu_a - \bar{\mu}_f = \epsilon_{amp} - b_{amp}, \tag{4.27}$$

where

$$b_{amp} = \bar{\mu}_f - \bar{\mu}_t \tag{4.28}$$

is the error in the prediction of the "true" mean precipitation $\bar{\mu}_t$ and

$$\epsilon_{amp} = \mu_a - \bar{\mu}_t. \tag{4.29}$$

Similarly, the ensemble members and the verification data are assumed to be from the same distribution and $\bar{\mu}_f$ is assumed to be an unbiased estimate of the mean of the distribution from which $\mu_f^k, k = 1, 2, ..., K'$ are drawn. Under these assumptions, for a sample size M, the expected value

$$E\left[\frac{1}{M} \sum_{m=1}^{M} \bar{\mu}_f^m\right] \tag{4.30}$$

of the sample mean of the predicted mean is equal to the explected value

$$E\left[\frac{1}{M} \sum_{m=1}^{M} \mu_a^m\right] \tag{4.31}$$

of the sample mean of the verification data.

Because μ_a is a realization of the true areal mean of the precipitation, μ_t,

$$E[\epsilon_{amp}] = E\left[\mu_a - \bar{\mu}_t\right] = 0. \tag{4.32}$$

Thus taking the expected value of Eq. (4.27) yields

$$E\left[\bar{\mu}_f - \bar{\mu}_t\right] = E[b_{amp}], \tag{4.33}$$

which leads to

$$\frac{1}{M}\sum_{m=1}^{M}\left(\mu_a^m - \bar{\mu}_f^m\right) \approx \frac{1}{M}\sum_{m=1}^{M} b_{amp}^m. \tag{4.34}$$

for a sufficiently large M. Therefore, for a sample of M forecasts, the necessary condition to be verifed for $b_{amp} = 0$ is

$$\frac{1}{M}\sum_{m=1}^{M}\left(\mu_a^m - \bar{\mu}_f^m\right) \approx 0. \tag{4.35}$$

By analogy to Eq. (4.18), the standard deviation (spread) of the ensemble of the areal mean precipitation,

$$\Sigma_{amp} = \left[\frac{1}{K\prime - 1}\sum_{k=1}^{K\prime}\left(\mu_f^k - \bar{\mu}_f\right)^2\right]^{1/2}, \tag{4.36}$$

is a prediction of the root-mean-square

$$\Delta_{amp} = \left[\frac{K\prime}{K\prime - 1}\left(\mu_a - \bar{\mu}_f\right)^2\right]^{1/2} \tag{4.37}$$

of the uncertainty in the ensemble mean of the areal mean precipitation. This leads to the verifiable necessary condition

$$\sigma_{amp} = \delta_{amp}, \tag{4.38}$$

where

$$\sigma_{amp} = \left[\frac{1}{M}\sum_{m=1}^{M}\left(\Sigma_{amp}^{m}\right)^2\right]^{1/2}, \tag{4.39}$$

$$\delta_{amp} = \left[\frac{1}{M}\sum_{m=1}^{M}\left[\Delta_{amp}^{m}\right)^2\right]^{1/2}.$$

4.4 Structure Forecasts

The two components of ASSIM that measure the structure error are the variance distortion D_{var} and the loss of correlation D_{cor}. Separate performance measures of the ensemble are derived for the two components of the structure error.

The derivation of the performance measure used for the variance in this study are simple. They are based on the recognition that for a properly designed ensemble and a good forecast model, the spatial variability of any atmospheric field should be the same in all forecast members and the analysis. This requirement leads to the verifiable condition

$$v_f = v_a, \tag{4.40}$$

where

$$v_f = \frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{K'(m)}\sum_{k=1}^{K'(m)}\left(\sigma_f^{mk}\right)^2\right], \tag{4.41}$$

and

$$v_a = \frac{1}{M}\sum_{m=1}^{M}\left(\sigma_a^{m}\right)^2. \tag{4.42}$$

In Eq. (4.41), σ_f^{mk} is the standard deviation σ_f for the k-th ensemble member of the m-th forecast, and in Eq. (4.42), σ_a^m is the standard deviation σ_a for the m-th analysis.

The derivation of the performance measure for the correlation is slightly more involved. One source of complexity is that the correlation between a forecast and an analysis

feature is affected by both the location error and the similarity between the two features. Because the goal here is to derive a measure for the performance of the ensemble in representing the uncertainty only in the similarity, but not the location, of the two features, all forecast members have to be first corrected for the location error. Let P_{fc}^k be the precipitation field of the k-th ensemble member after the correction of the location error, which can be done by shifting the pixels of P_f^k by $d\mathbf{X}^k$. In addition, let

$$\bar{P}_{fc} = \frac{1}{K'} \sum_{k=1}^{K'} P_{fc}^k. \tag{4.43}$$

If the ensemble samples the (true) probability distribution of the precipitation,

$$E\left[\text{cor}\left(P_{fc}^k, \bar{P}_{fc}\right)\right] = E\left[\text{cor}\left(P_a, \bar{P}_{fc}\right)\right], \tag{4.44}$$

where $\text{cor}\left(P_x, P_y\right)$ is the correlation between the precipitation fields P_x and P_y. Equation (4.44) leads to the verifiable condition

$$c_f \approx c_a, \tag{4.45}$$

where

$$c_f = \frac{1}{M} \sum_{m=1}^{M} \left[\frac{1}{K'(m)} \sum_{k=1}^{K'(m)} \text{cor}\left(P_{fc}^{mk}, \bar{P}_{fc}^m\right) \right] \tag{4.46}$$

$$c_a = \frac{1}{M} \sum_{m=1}^{M} \text{cor}\left(P_a^m, \bar{P}_{fc}^m\right). \tag{4.47}$$

5. FORECAST AND VERIFICATION DATA

5.1 Winter Storm Case Description

This section gives a detailed description of the winter storm cases used in the calculations. These cases includes 32 United States named winter storms from the 2014/2015 and 2015/2016 season. The date, verification time, domian, and sample size for each storm is listed in Table 5.1. The verification variable is the 6 h acumulated precipitation starting at the verification time in Table 5.1. For example, verification time 0000 h UTC and date 11/10/2014 indicates the verification of forecasts (of various lead times) of the 6 h accumulation between 0000 UTC and 0600 UTC, November 10, 2014. The verification times and verification domains are chosen such that a well-defined precipitation system is present in each case. The size of the sample of verification cases is 133.

5.2 Forecast Data

The forecast used for our investigation are $0.5^\circ \times 0.5^\circ$ resolution global 20-member ensemble forecasts of the National Centers for Environmental Prediction (NCEP) that are archived in the THORPEX Interactive Grand Global Ensemble (TIGGE) dataset. This dataset is a collection of global ensemble forecasts from the major numerical weather prediction (NWP) centers of the world (Bougeault et al., 2010; Swinbank et al., 2016). In this data archive, the NCEP ensemble data are available for the first 16 forecast days.

5.3 Verification Data

The verification dataset is a combination of the Stage IV analysis product of NCEP and operational 0-6 h precipitation forecasts from the European Centre for Medium Range Forecasts (ECMWF). (The latter data is also available from the TIGGE dataset.) The Stage IV precipitation analysis is based on radar and gauge observations of the precipitation over

Table 5.1: Description of the 32 United States named winter storms.

Winter Storm	Date (MM/DD/YYYY, UTC)	Verif. Time (hhmm, UTC)	Domain (North/West/South/East)	Sample Size
Astro	11/10/2014	1200	75.5°N/134°W/36°N/78.5°W	1
	11/11/2014	1200	75.5°N/134°W/36°N/78.5°W	1
Bozeman	11/16/2014	1200	62.5°N/100°W/23°N/44.5°W	1
	11/17/2014	0000,1200	62.5°N/100°W/23°N/44.5°W	2
Cato	11/26/2014	1200	57.5°N/105°W/18°N/49.5°W	1
Damon	12/09/2014	1200	57.5°N/108°W/18°N/52.5°W	1
	12/10/2014	0000	57.5°N/108°W/18°N/52.5°W	1
Westcoast	12/11/2014	0000	57.5°N/158°W/18°N/102.5°W	1
Hektor	01/12/2015	0000	57.5°N/108°W/18°N/52.5°W	1
Iola	01/24/2015	0000	57.5°N/108°W/18°N/52.5°W	1
Linus	02/02/2015	0000,1200	57.5°N/103°W/18°N/47.5°W	2
Octavia	02/17/2015	0000	54.5°N/111°W/15°N/47.5°W	1
Pandora	02/21/2015 to 02/22/2015	0000,1200	57.5°N/125°W/18°N/69.5°W	4
Remus	02/26/2015	0000	57.5°N/113°W/18°N/57.5°W	1
Sparta	03/02/2015	0000	57.5°N/110°W/18°N/54.5°W	1
	03/03/2015	1200	57.5°N/110°W/18°N/54.5°W	1
	03/04/2015 to 03/05/2015	0000,1200	57.5°N/110°W/18°N/54.5°W	4
Ajax	11/17/2015 to 11/18/2015	0000,1200	57.5°N/120°W/18°N/64.5°W	4
Cara	11/27/2015 to 11/28/2015	0000,1200	63.5°N/123°W/24°N/67.5°W	4
Delphi	11/30/2015 to 12/01/2015	0000,1200	63.5°N/125°W/24°N/69.5°W	4
Echo1	12/13/2015 to 12/15/2015	0000,1200	63.5°N/110°W/24°N/54.5°W	6
Echo2	12/16/2015	0000,1200	63.5°N/125°W/24°N/69.5°W	2
Ferus	12/23/2015	0000,1200	63.5°N/115°W/24°N/59.5°W	2
Goliath	12/27/2015 to 12/29/2015	0000,1200	63.5°N/115°W/24°N/59.5°W	6
Hera	01/10/2016 to 01/11/2016	0000,1200	59.5°N/113°W/20°N/57.5°W	4
Jonas	01/22/2016	0000,1200	54.5°N/113°W/15°N/57.5°W	2
	01/23/2016	0000	54.5°N/113°W/15°N/57.5°W	1
Lexi	02/04/2016 to 02/05/2016	0000,1200	64.5°N/110°W/25°N/54.5°W	4
Mars	02/07/2016	0000	59.5°N/105°W/20°N/49.5°W	1
Olympia	02/16/2016	0000,1200	59.5°N/110°W/20°N/54.5°W	2
Petros	02/23/2016 to 02/24/2016	0000,1200	59.5°N/110°W/20°N/54.5°W	4
	02/25/2016	0000	59.5°N/110°W/20°N/54.5°W	1
Quo	03/02/2016	0000	59.5°N/110°W/20°N/54.5°W	1
Selene	03/24/2016	0000,1200	69.5°N/120°W/30°N/64.5°W	2
Troy	03/31/2016	0000,1200	69.5°N/120°W/30°N/64.5°W	2
Vexo	04/16/2016 to 04/18/2016	0000,1200	66.5°N/120°W/27°N/64.5°W	6
TOTAL				83

the United States (Lin and Mitchell, 2005). The analyses are available as hourly, 6 hourly, and daily rainfall accumulations for approximately $4\ km \times 4\ km$ pixels. Since most winter storms have precipitation systems that extend over the ocean, where Stage IV data is not available, the Stage IV data is combined with the $0.5^\circ \times 0.5^\circ$ resolution ECMWF forecasts to generate a composite precipitation field that covers the entire verification region for each storm.

The composite precipitation fields are obtained by first upscaling the $4\ km \times 4\ km$ Stage IV analyses to the resolution of the ECMWF forecasts, which is identical to the resolution of the NCEP ensemble forecasts to be verified. The finer resolution Stage IV analyses are interpolated onto the $0.5^\circ \times 0.5^\circ$ grid by an area-mean interpolation to conserve total precipitation. The interpolated Stage IV analysis data are then augmented by the ECMWF forecast data at the locations where no Stage IV data is available. To ensure that the augmented fields reflect the spatial statistical properties of the Stage IV data, the ECMWF data has to be first calibrated. This calibration is done such that for the locations where both Stage IV analyses data and ECMWF forecast data are available, the mean and the standard deviation of the logarithm of the values in the two data sets are the same.

The effects of calibration of the ECMWF forecasts on the composite (ST4-ECMWF) precipitation fields are illustrated by two examples. Figure 5.1 shows the ECMWF forecast, regridded Stage IV analysis, and ST4-ECMWF for the 6 h precipitation accumulation for Winter Storm Frona starting at 0000 UTC, December 29, 2014. Since the ECMWF precipitation over land is weaker than the one in the regridded Stage IV analysis, the calibration of the ECMWF forecast increases the precipitation amplitude and variance over the entire verification domain. Figure 5.2 is the same as Fig. 5.1, except for Winter Storm Quo starting at 0000 UTC, March 02, 2016. As in the previous example, the main forecast precipitation system located in the middle of the verification domain by ECMWF matches the precipitation system of the Stage IV analysis. However, the magnitude of the forecast

system, as inferred by the comparison of the two fields, is underestimated.

Figure 5.3 shows the histogram of the ratios

$$r^m = \frac{\mu^m_{st4}}{\mu^m_{st4ec}}, m = 1, ..., M \tag{5.1}$$

where μ^m_{st4} is the μ_a for the regridded Stage IV total precipitation of the m-th winter storm, and μ^m_{st4ec} is μ_a for the composite precipitation field of the m-th winter storm. The larger the ratio r^m, the larger the weight of the Stage IV analysis in the composite precipitation field. The results of Chapter 6 are based on those cases for which $r^m > 0.3$. The number of these cases is 83, and their description is included in Table 5.1.

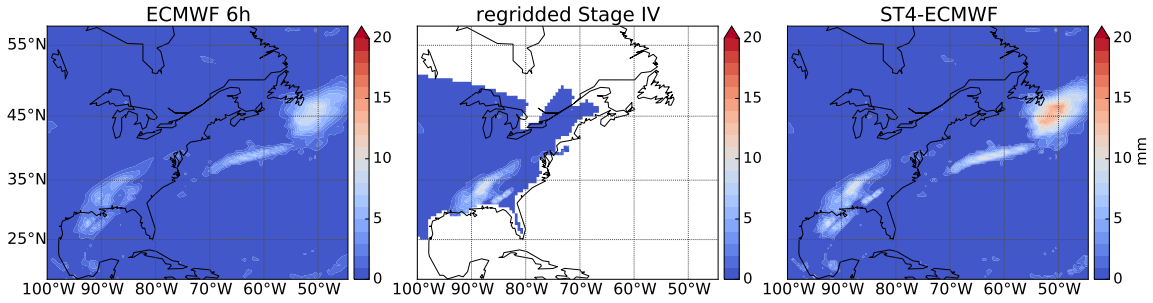

Figure 5.1: (left) ECMWF 6 h accumulated precipitation forecast, (middle) regridded Stage IV 6 h precipitation analysis, and (right) combined ST4-ECMWF precipitation field for Winter Storm Frona starting at 0000 UTC, December 29, 2014.

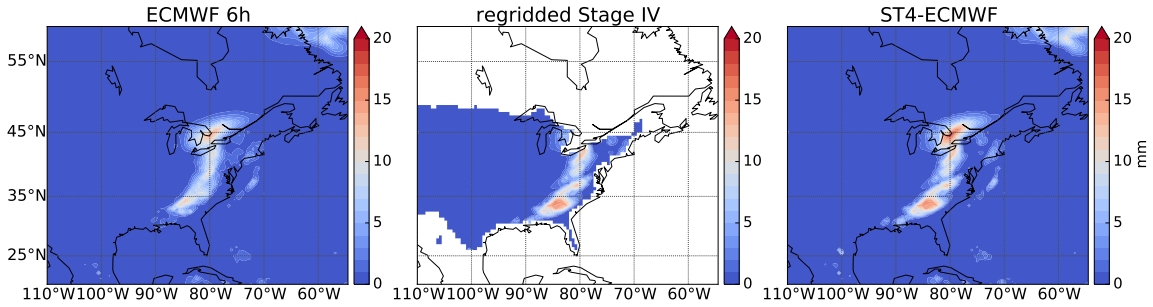

Figure 5.2: Same as Fig. 5.1, except for Winter Storm Quo starting at 0000 UTC, March 02, 2016.

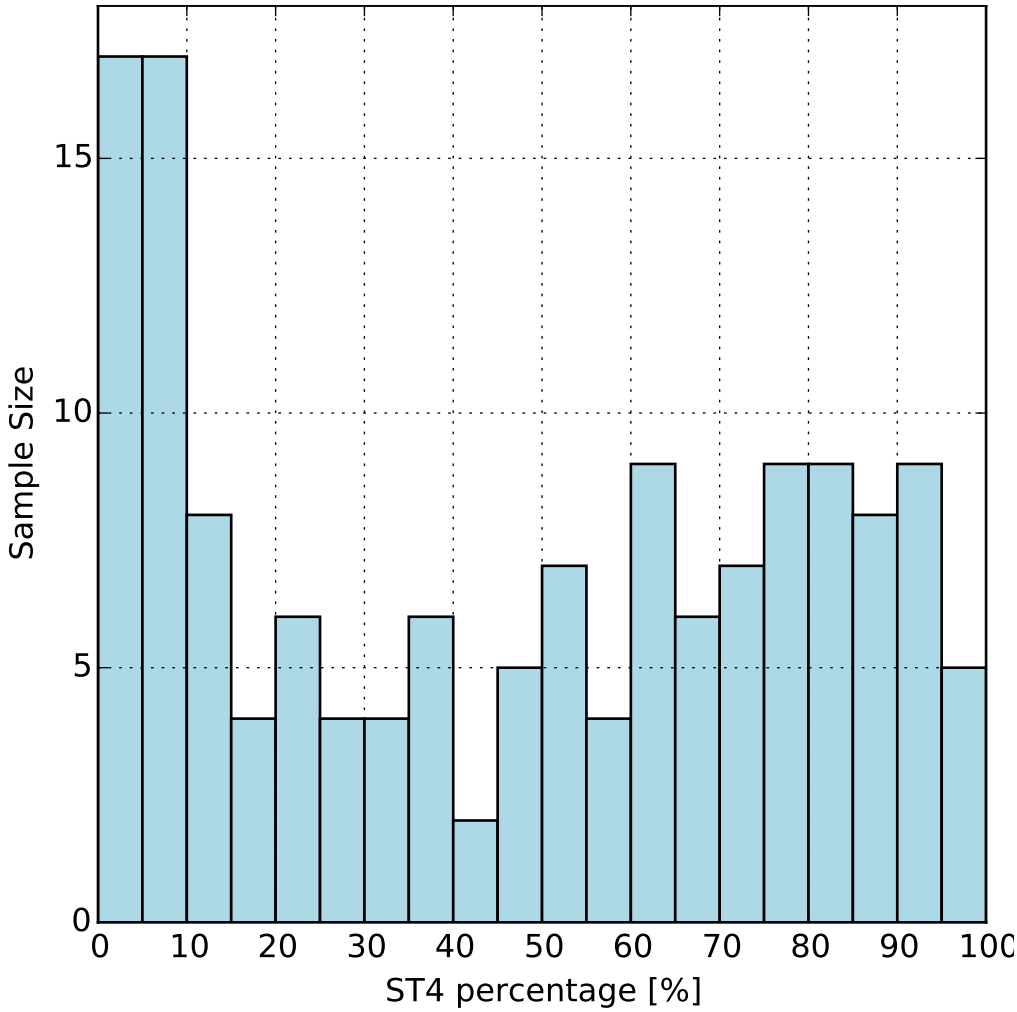

Figure 5.3: Histogram of the percentages of the Stage IV total precipitation over the total precipitation of the composite ST4-ECMWF precipitation of the 133 samples.

6. VERIFICATION RESULTS AND DISCUSSIONS

This chapter illustrates the verification results for the NCEP ensemble precipitation forecasts. The morphing based verification methodology presented in Chapter 3 is applied to the NCEP ensembles of 83 winter storm cases described in Chapter 5. Uncertainty metrics introduced in Chapter 4 are computed based on the morphing results. The parameters of the morphing based techniques used in the calculations are the same as those that were used in the examples illustrated in Chapter 3.

6.1 Sensitivity of the Average K' to the Choice of the ASSIM Criteria

After the morphing based technique is applied to all the ensemble members of the m-th ensemble, the value of $K'(m)$ is determined. $K'(m) \leq 20$ is the number of ensemble members that successfully predict the verifying event. As was discussed in Section 4.1, an ensemble member successfully predicts the verifying event if the similarity between the shifted forecast and the verifying analysis, measured by ASSIM, is greater than a predefined threshold δ. This section compares the evolution of the average value of K' with lead time, using different values of δ. Since only ensembles with 2 or more ensemble members that successfully predict the verifying event are used in the calculations of uncertainty metrics, the average is computed over all samples with $K'(m) \geq 2$.

A common feature of the evolutions of K' for different values of δ is that it decreases as the lead time increases (top panel of Fig. 6.1). For $\delta = 0.6$ (purple), $\delta = 0.7$ (green), and $\delta = 0.8$ (red), the average K' starts at 20 and saturates approximately after 10 forecast days at a level of 8, 5, and 3, respectively. For the largest δ value we tested ($\delta = 0.9$), the evolution of the average K' started at less than 20 and quickly decreased to zero. The evolution of the percentage of the ensembles (samples) in which $K'(m) \geq 2$ for the different values of δ are also shown in the bottom panel of Fig. 6.1. Unlike for the average

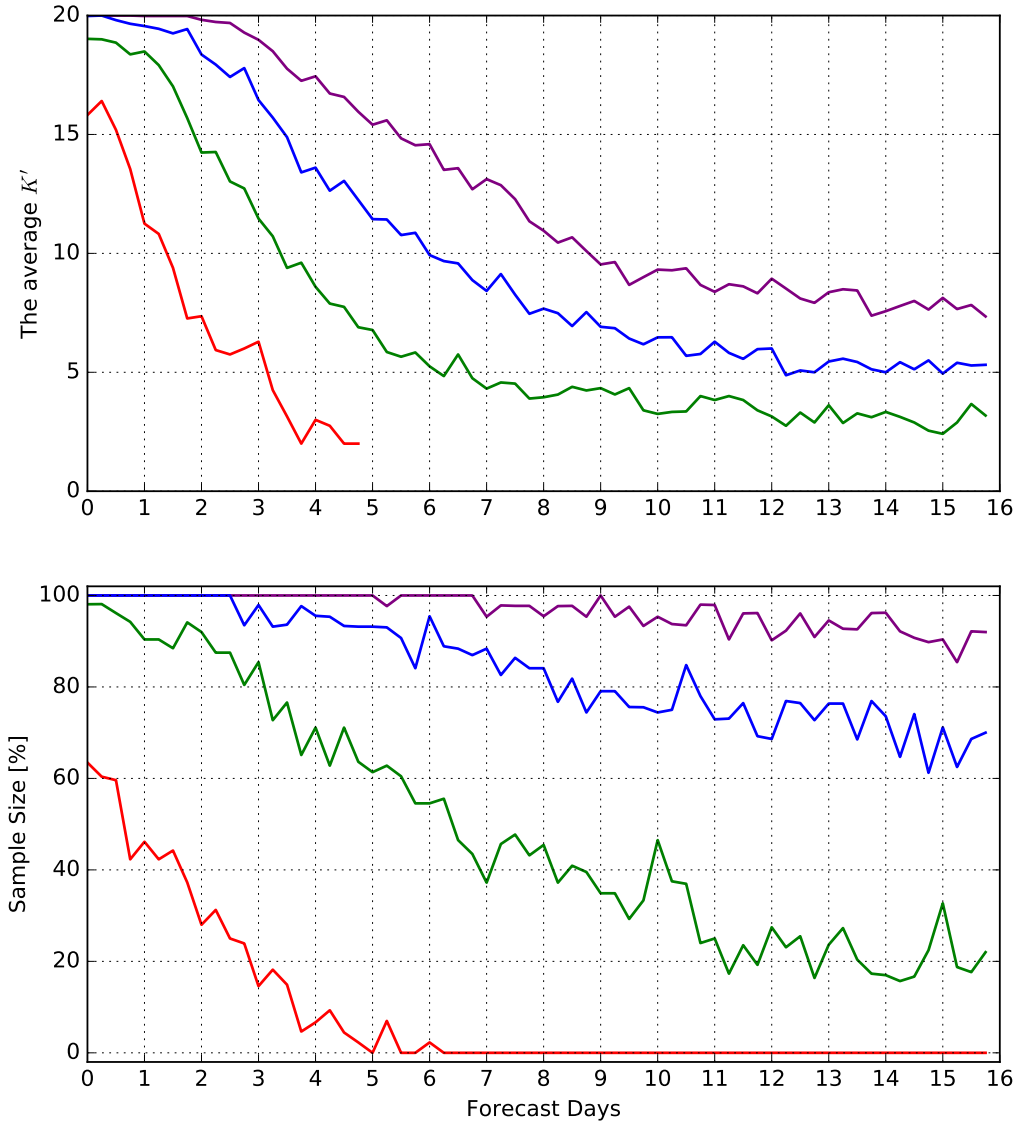

Figure 6.1: Evolutions of the average K' and percentages of ensembles in which $K' \geq 2$ with different δ values: (purple) $\delta = 0.6$, (blue) $\delta = 0.7$, (green) $\delta = 0.8$, and (red) $\delta = 0.9$.

K', the only curve that shows a clear saturation is the $\delta = 0.8$ one.

6.2 Uncertainty Results for $\delta = 0.8$

6.2.1 Location Uncertainty

Figure 6.2 illustrate the bahavior of the evolutions of σ_{pos} and δ_{pos} calculated using Eq. (4.21) and Eq. (4.19). As shown in this figure, the ensemble spread of the magnitude of the location vectors matches very well with the ensemble mean error of the magnitude of the location vectors at shorter forecast lead days. This indicates that the NCEP ensemble forecasts are highly skillful in predicting the uncertainty of the locations of the storms. After the 5^{th} forecast days, the ensemble spread deviates slightly from the ensemble mean error, but they saturate at similar levels in the end.

6.2.2 Amplitude Uncertainty

Figure 6.3 illustrate the bahavior of the evolutions of σ_{amp} and δ_{amp} calculated using Eqs. (4.36), (4.37), and (4.39). The ensemble greatly underestimates the uncertainty of the precipitation amplitude at the shorter forecast times. Even though the difference between σ_{amp} and δ_{amp} quickly narrows with increasing lead time, there is no clear match before they staruates at similar levels at about forecast day 5.5-6. This indicates that the NCEP ensemble forecasts have limited capability in predicting the uncertainty in the mean magnitudes of the storms at short forecast times.

6.2.3 Structure Uncertainty

6.2.3.1 Variance

Figure 6.4 shows the results for the variance component of the structure error. The results indicate that the variance of the precipitation field is lower for the forecasts than the analysis at the shorter forecast times. In other words, at the short forecast times the forecast fields are spatially smoother than the analysis field. The diminishing difference between

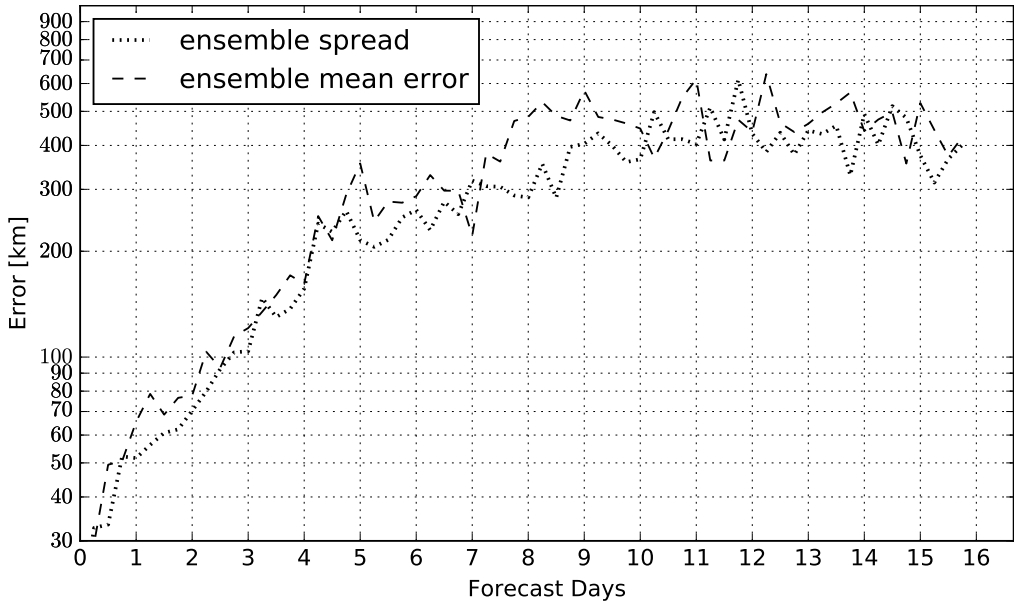

Figure 6.2: Evolution of σ_{pos} (ensemble spread) and δ_{pos} (ensemble mean error).

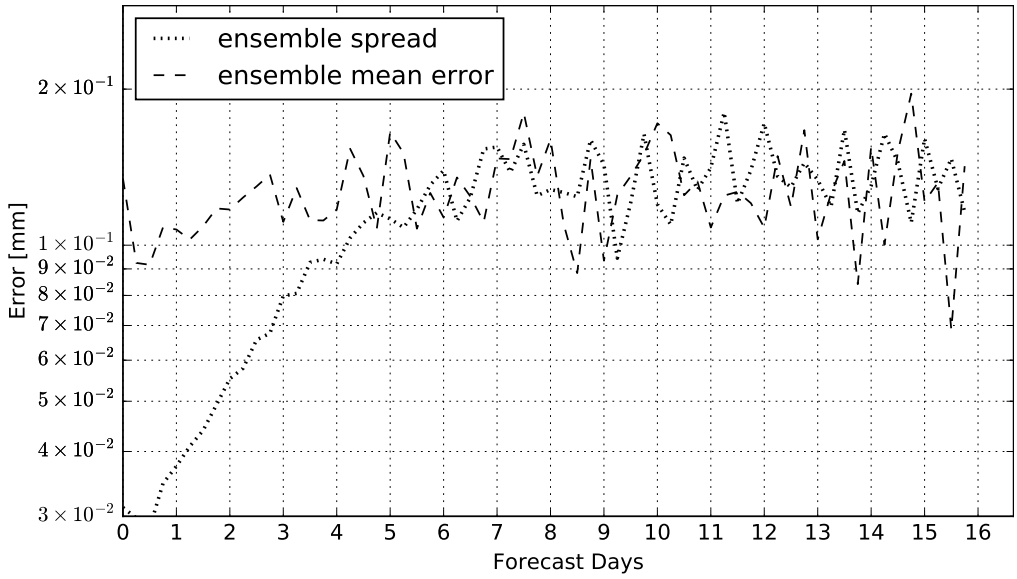

Figure 6.3: Evolution of σ_{amp} (ensemble spread) and δ_{amp} (ensemble mean error).

the smoothness of the forecast and analysis fields with increasing forecast time suggests that the variance deficiency of the short-term forecasts is due to a variance deficiency of the initial conditions (analyses) rather than deficiencies of the forecast model.

6.2.3.2 Correlation

The results for the correlation component of the structure uncertainty are shown in Figure 6.5. For the interpretation of this figure, it should be recalled that ensemble averaging is a nonlinear filter: it removes spatial details of the forecasts that are not predictable due to the chaotic nonlinear dynamics of the atmosphere. If the initial condition uncertainty was infinitesimally small, and the ensemble correctly reflected that property, both c_a and c_f would have a value of 1 at initial time (day 0). (In essence, the ensemble members would be nearly identical to the analysis and each other.) The values of c_a and c_f much smaller than 1 indicate that reality is far from the ideal situation. While the ensemble does a good job with capturing the large initial condition uncertainty in the point-wise structure of the precipitation field, it still slightly underestimates the uncertainty.

As forecast time increases, both c_a and c_f must decrease due to the increasing number of unpredictable details. It is worth noting that the difference between the two curves decreases with forecast time, indicating that the ensemble provides an increasingly better representation of the uncertainty in the point-wise structure of the precipitation field.

6.3 Sensitivity of the Uncertainty Results to the Choice of the ASSIM Criteria

In this section, the uncertainty results calculated using 4 different values of δ (0.6, 0.7, 0.8, and 0.9) are compared and discussed.

The results for the location uncertainty are shown in Figure 6.6. The evolutions of σ_{pos} (ensemble spread) and δ_{pos} (ensemble mean error) are very similar, including the saturation levels (approximately $450\ km$) except for the ones in the bottom panel ($\delta = 0.9$). For the interpretation of the results for $\sigma = 0.9$, recall from Section 6.1 that the sample size drops

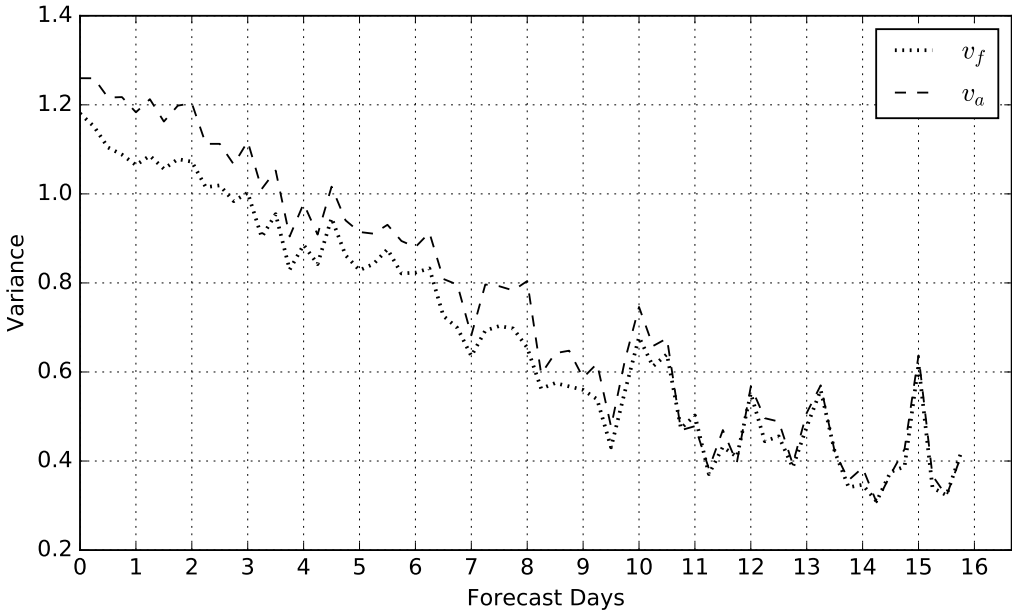

Figure 6.4: Evolution of v_f and v_a.

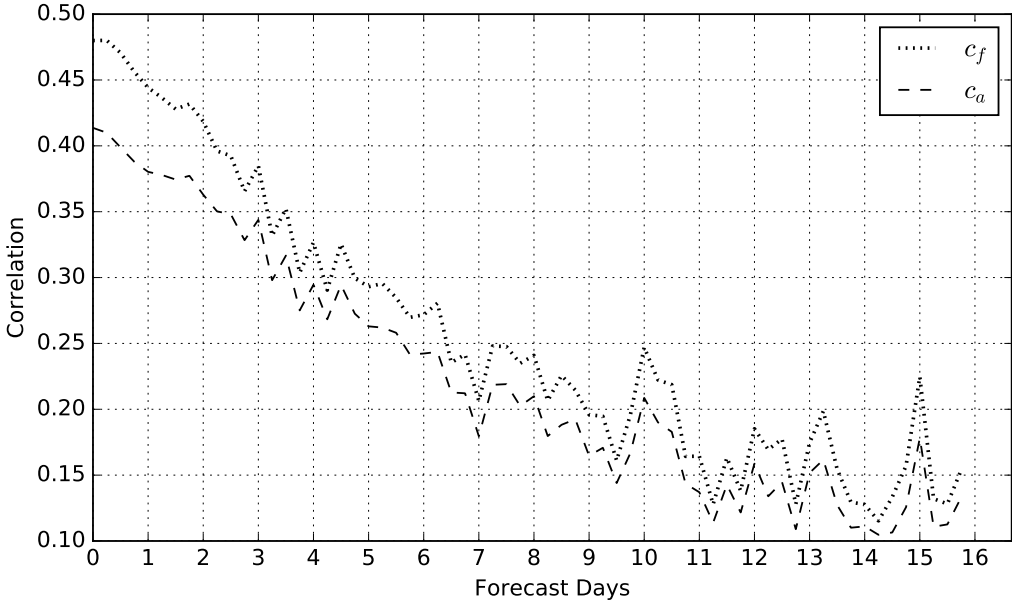

Figure 6.5: Evolution of c_f and c_a.

to zero after 5 forecast days.

The results for the amplitude uncertainty are shown in Figure 6.7. The first three panels of the amplitude uncertainty results show consistently similar patterns to those discussed in Section 6.2, indicating that the amplitude uncertainty results are very robust and not sensitive to the choice of the ASSIM criteria. The saturation levels, however, do vary with the value of ASSIM criteria. The larger the value of δ, the lower the level of saturation.

The results for the variance part of the structure uncertainty (Fig. 6.8) show very different evolution patterns for the different values of δ. A common feature of the behavior for all the panels is that the spatial variance of the precipitation field is always lower for the forecasts than the analysis. Figure 6.9 shows the curves of the differences between the two variances, $v_a - v_f$. Unlike for $\delta = 0.8$, the difference between the forecast and the analysis variance does not decrease until after forecast day 14 and 10 for $\delta = 0.6$ and $\delta = 0.7$, respectively. The difference for the $\delta = 0.6$ case even shows an increasing trend in shorter forecast times. This indicates that a better agreement between the forecasts and the analysis variance at the longer foreacst times can only be reached by a proper selection of ensemble members by the ASSIM criterion.

The results for the correlation part of the structure uncertainty are shown in Fig. 6.10. For shorter forecast times, a common feature of the behavior for all 4 panels is that the ensemble underestimates the uncertainty in the point-wise structure of the precipitation field. The underestimation is most dominant for the first 2 days and weakens as forecast time increases. Another common feature is the decreasing trend of the values of both c_f and c_a despite their differences due to the increasing number of unpredictable details in the point-wise structure of the precipitation forecast. A distinct feature of the curves in the top two panels with $\delta = 0.6$ and 0.7 is that the difference between the two curves decreases at shorter forecast times and reaches a minimum at forecast days 3-7. Beyond forecast day 7, the difference between the two curves starts increasing.

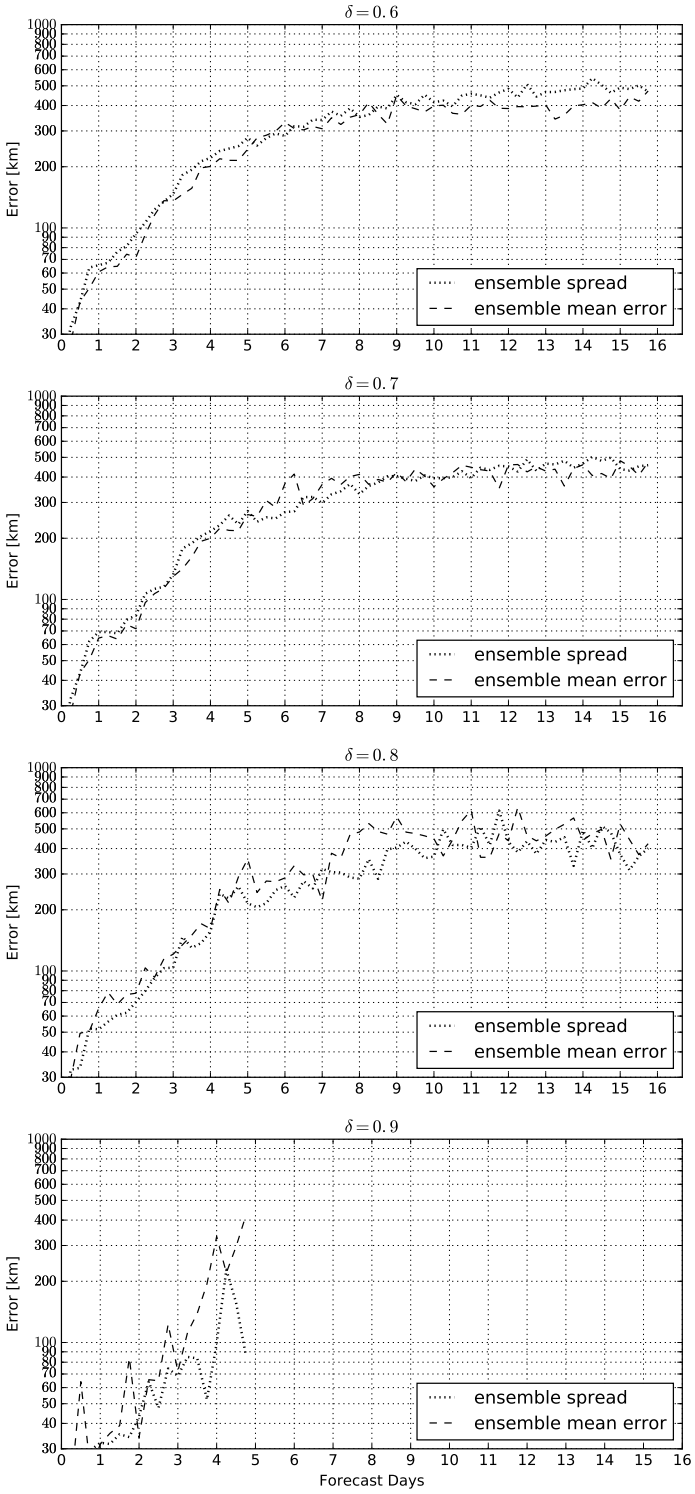

Figure 6.6: Evolutions of σ_{pos} and δ_{pos} for $\delta = 0.6, 0.7, 0.8$, and 0.9.

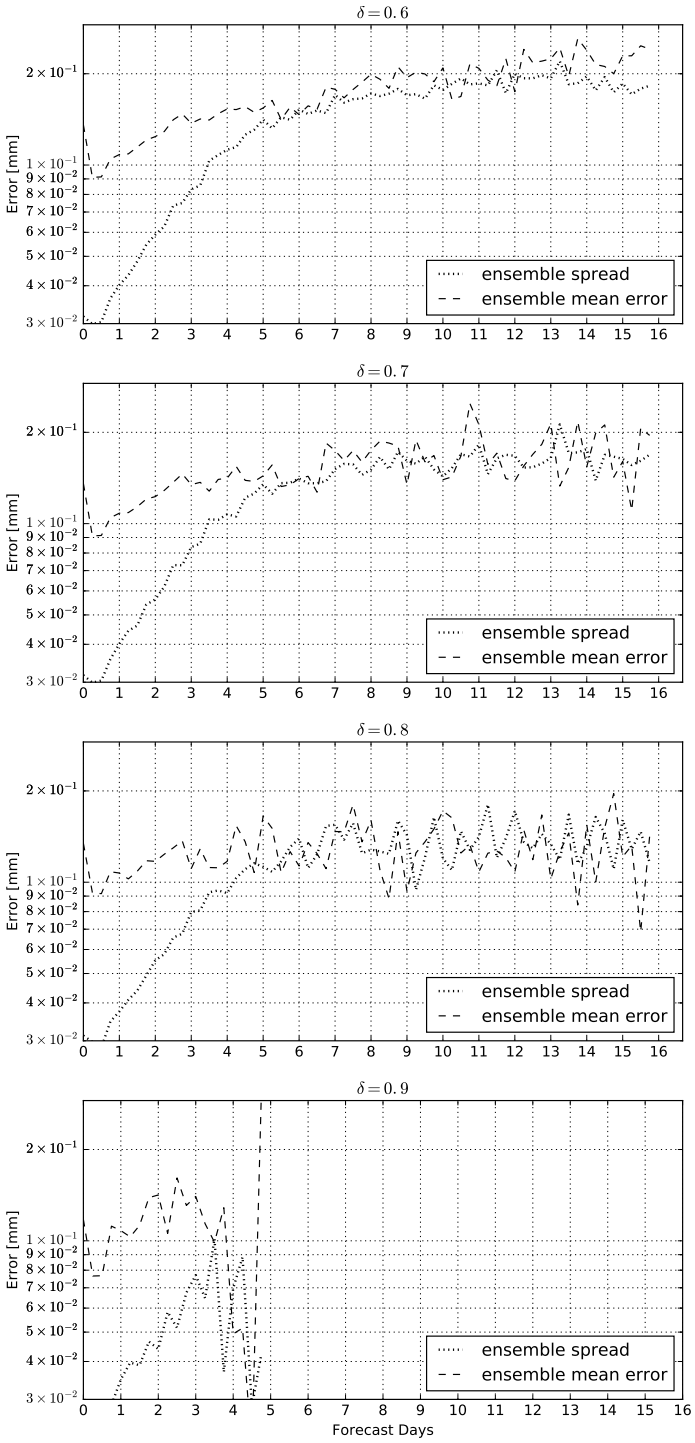

Figure 6.7: Evolutions of σ_{amp} and δ_{amp} for $\delta = 0.6$, 0.7, 0.8, and 0.9.

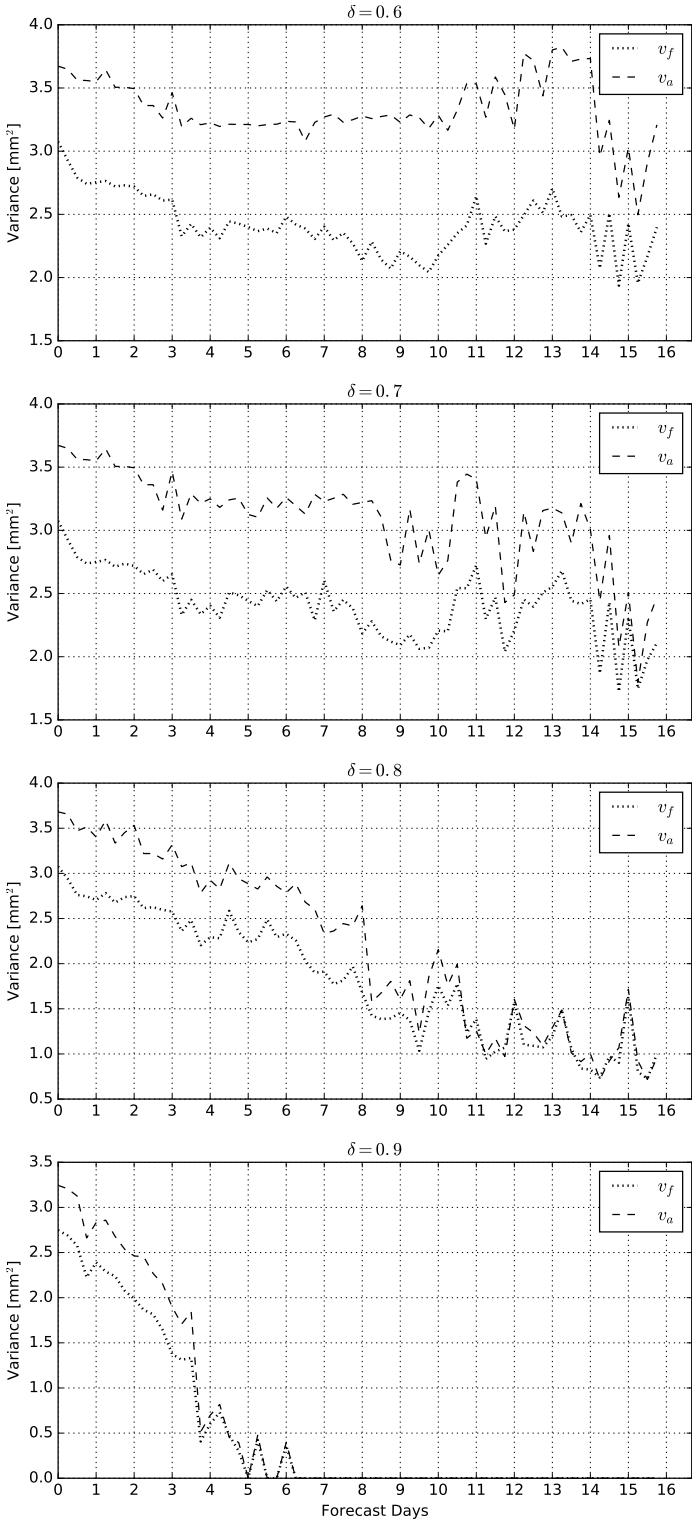

Figure 6.8: Evolutions of v_f and v_a for $\delta = 0.6$, 0.7, 0.8, and 0.9.

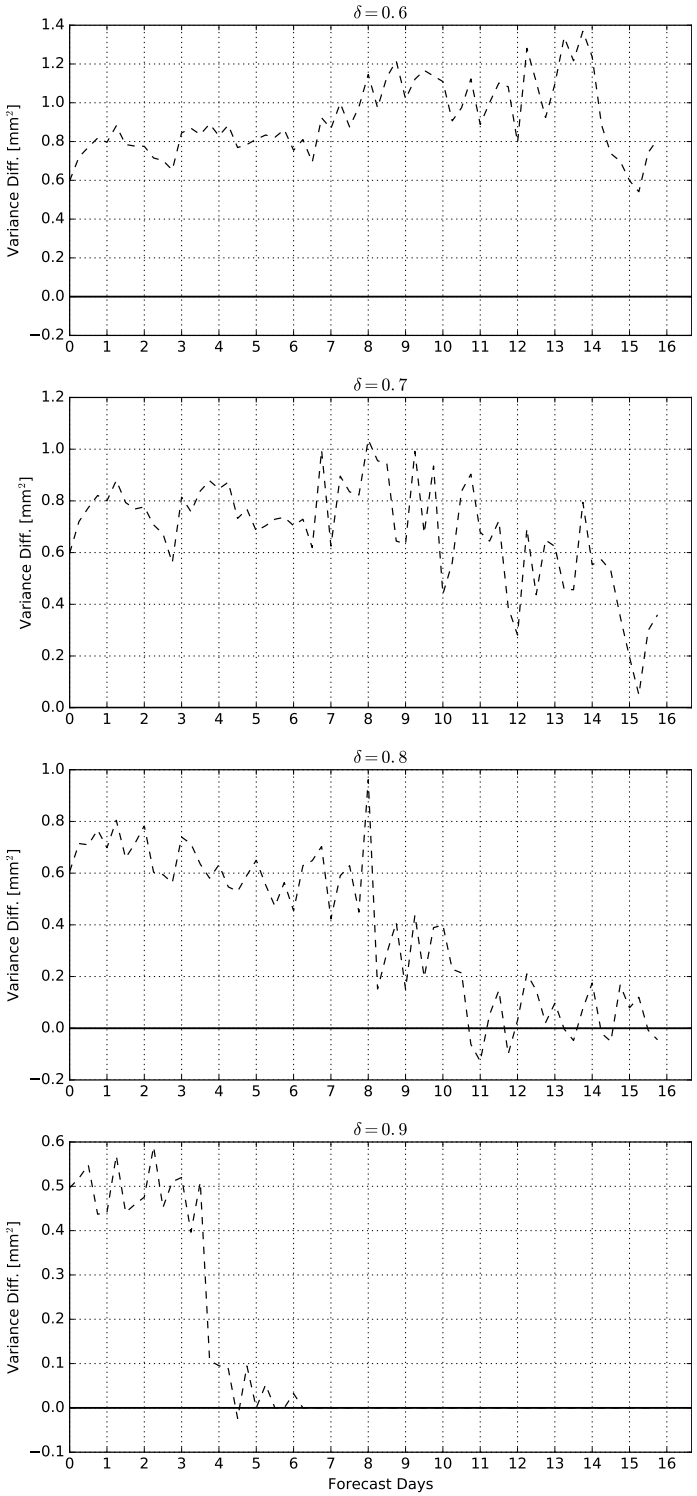

Figure 6.9: Evolutions of $v_a - v_f$ for $\delta = 0.6, 0.7, 0.8$, and 0.9.

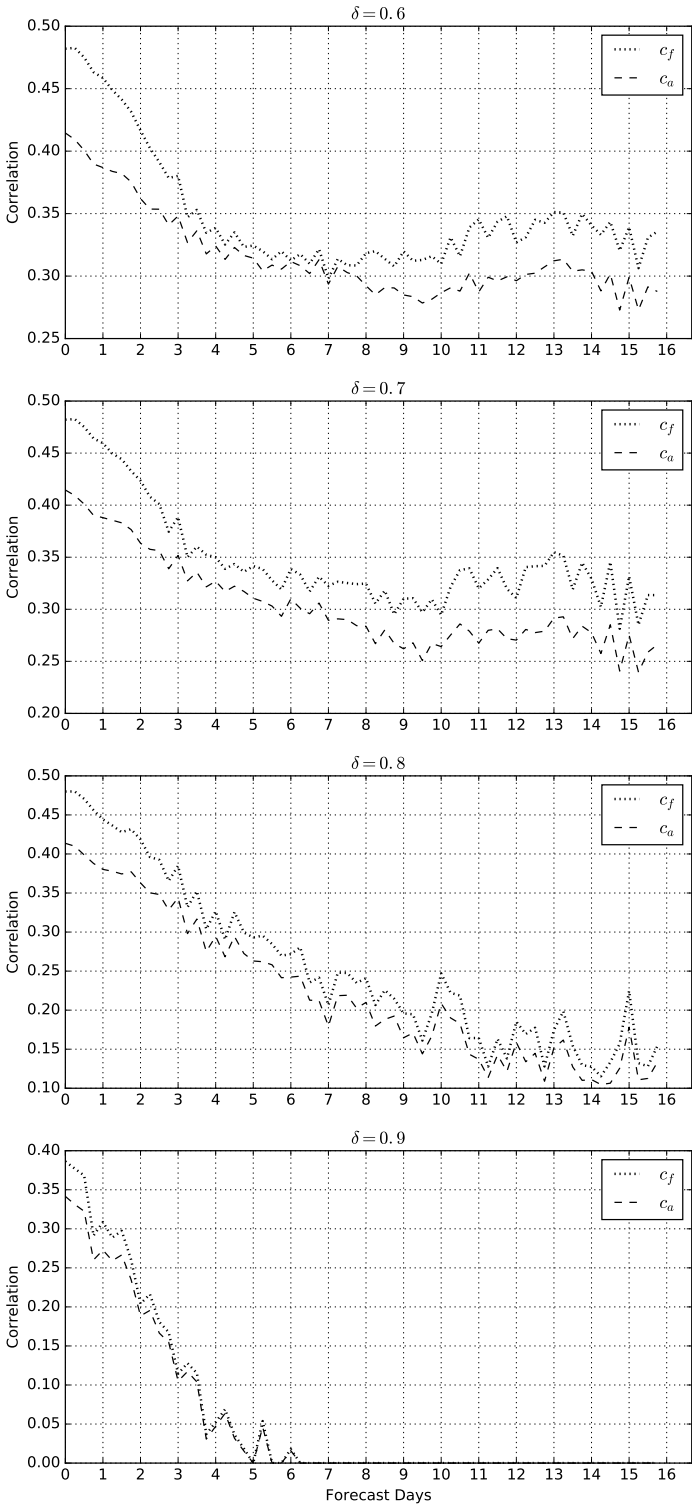

Figure 6.10: Evolutions of c_f and c_a for $\delta = 0.6, 0.7, 0.8$, and 0.9.

7. CONCLUSIONS

A new morphing based technique, which is an improved version of the technique of HS16, was proposed for the verification of deterministic precipitation forecasts. The new technique can accurately estimates of the components of the precipitation forecast error for a significantly longer range of the location error, and is computationally more efficient, than the original technique.

Motivated by the UQI (Wang and Bovik, 2002) and the SSIM (Wang and Bovik, 2009; Wang et al., 2004) image quality indexes, a new index, called ASSIM, was also introduced. ASSIM was used to measure the similarity between precipitation fields in the morphing algorithm, and to quantify amplitude and structure error of the forecasts. When ASSIM was used to measure the forecast quality, it provided forecast scores that were in a better agreement than the other tested objective scores with the scores by a group of precipitation forecast experts. The correspondence between the ASSIM based scores and the expert scores further improved when the forecasts were first corrected for location error.

A set of diagnostic equations that take advantage of the morphing based technique was also derived for the verification of ensemble precipitation forecasts. An application of these diagnostic equations to the operational global NCEP ensemble forecasts of the named winters storms of the 20015/2016 winter storm season provided new insights into the predictability of winter storms and the skill of the operational ensemble forecasts in predicting that predictability.

One measure of the predictability of a storm is the number of ensemble members that predict the precipitation system of the storm. Whether an ensemble member predicted a particular storm or not was determined with the help of ASSIM: an ensemble member was deemed to capture a storm successfully, if ASSIM for the forecast corrected for location

error and the verifying analysis was larger than a prescribed value δ. It was found that δ had to be smaller than or equal to 0.8, because for larger values of δ, a significant number of ensemble members were found not capture the storm even at initial time. For values of δ that satisfied the criterion $\delta \leq 0.8$, the number of ensemble members that were found to capture the storms saturated at a level larger than zero. This behavior indicates that once the saturation level is reached, a storm in the forecast was found similar to a random (unrelated) storm in the verifying analysis. Thus, the predictability limit can be defined by the the forecast time at which the saturation level is reached. The predictability limits for $\delta = 0.8$, $\delta = 0.7$, and $\delta = 0.6$ are 10, 12, and 14 days, respectively.

The results for the predictability of the location and the total precipitation amount were found to be highly robust to the choice of δ: the location error grew exponentially for 3-4 days and saturated at forecast time 8-9 days. In other words, the error in the prediction of the location of the storm grows rapidly with forecast time, making the location completely unpredictable beyond 8-9 days on average. A much welcome finding was that the ensemble provided an on average highly accurate quantitative prediction of the uncertainty of the location. Because the uncertainty in the location is the dominant factor in determining the likelihood of the precipitation at a specific time and location, forecast users can have a high level of trust in the ensemble based predictions of the probability of precipitation. It should be emphasized, however, that this result does not imply that forecast users should also trust the predicted probability of specific amounts (threshold values) of the precipitation. In fact, the results showed that at forecast times shorter than 7-8 days, the ensemble grossly underestimates the uncertainty of the precipitation amount.

The good performance of the ensemble in predicting the location uncertainty indicates that the technique used for the generation of the ensemble initial conditions does a good job in sampling the analysis uncertainty in the location of the storm. It also indicates that the model does a good job in describing the spatiotemporal evolution of that uncertainty.

Qualitatively, this result is not surprising, because ensemble generation techniques tend to focus on capturing the synoptic scale analysis uncertainties, which play the key role in determining the location uncertainty of the winter storms. In addition, numerical weather prediction models are known to be skillful in describing the synoptic scale dynamics. Nevertheless, *the almost perfect quantitative performance of the ensemble in predicting the location uncertainty is an unexpected and highly non-trivial result that could not have been detected without a robust technique for the estimation of the location error.*

The poor performance of the ensemble in predicting the uncertainty of the amplitude (total precipitation) at the short forecast times indicates that the technique used for the generation of the ensemble initial conditions is highly inefficient in capturing the related sources of uncertainty. The fact that at the longer forecast times (beyond about 5-6 days) the ensemble correctly predicted the saturation level of the uncertainty also suggests that the poor short term performance of the ensemble was due to shortcomings of the initial condition generation technique rather than shortcomings of the forecast model. While this behavior is not unexpected in general, the extent of the problem at the shortest forecast times is surprising. It points to the importance of further improving the techniques for the generation of ensemble initial conditions and suggests that short-term ensemble forecasts of the precipitation could greatly benefit from statistical post-processing.

The decrease of the spatial variance of the verifying analyses with increasing forecast time indicates that precipitation systems that have smoother spatial structures are easier to predict for a longer time. While this effect made the diagnostics for the structure uncertainty non-robust to the choice of δ, they still clearly indicated that the ensemble had a tendency to underestimate the structure uncertainty.

REFERENCES

Ahijevych, D., E. Gilleland, B. G. Brown, and E. E. Ebert, 2009: Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts. *Weather and Forecasting*, **24**, 1485–1497.

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, **9 (7)**, 1518–1530.

Anderson, J. L., 1997: The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Monthly Weather Review*, **125 (11)**, 2969–2983.

Bougeault, P., and Coauthors, 2010: The THORPEX interactive grand global ensemble. *Bulletin of the American Meteorological Society*, **91 (8)**, 1059–1072.

Brewster, K. A., 2003: Phase-correcting data assimilation and application to storm-scale numerical weather prediction. part i: Method description and simulation testing. *Monthly Weather Review*, **131 (3)**, 480–492.

Candille, G., C. Côté, P. Houtekamer, and G. Pellerin, 2007: Verification of an ensemble prediction system against observations. *Monthly Weather Review*, **135 (7)**, 2688–2699.

Casati, B., and Coauthors, 2008: Forecast verification: current status and future directions. *Meteorological Applications*, **15 (1)**, 3–18.

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. part I: Methodology and application to mesoscale rain areas. *Monthly Weather Review*, **134 (7)**, 1772–1784.

Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmospheric Science Letters*, **5 (6)**, 110–117.

Ebert, E. E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Weather and Forecasting*, **24 (6)**, 1498–1510.

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrometeorology*, **239 (1)**, 179–202.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21 (6)**, 739–759.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, **24 (5)**, 1416–1430.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, **129 (3)**, 550–560.

Hamill, T. M., and S. J. Colucci, 1997: Verification of eta-rsm short-range ensemble forecasts. *Monthly Weather Review*, **125 (6)**, 1312–1327.

Han, F., and I. Szunyogh, 2016: A morphing-based technique for the verification of precipitation forecasts. *Monthly Weather Review*, **144 (1)**, 295–313.

Hoffman, R. N., Z. Liu, J.-F. Louis, and C. Grassoti, 1995: Distortion representation of forecast errors. *Monthly Weather Review*, **123 (9)**, 2758–2770.

Keil, C., and G. C. Craig, 2007: A displacement-based error measure applied in a regional ensemble forecasting system. *Monthly Weather Review*, **135 (9)**, 3248–3259.

Keil, C., and G. C. Craig, 2009: A displacement and amplitude score employing an optical flow technique. *Weather and Forecasting*, **24 (5)**, 1297–1308.

Leith, C., 1974: Theoretical skill of monte carlo forecasts. *Monthly Weather Review*, **102 (6)**, 409–418.

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: development and applications. *19th Conferefence of Hydrology, American Meteorological Society*, URL https://ams.confex.com/ams/pdfpapers/83847.pdf.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal

resolution produce more skillful forecasts? *Bulletin of the American Meteorological Society*, **83 (3)**, 407–430.

Mo, R., C. Ye, and P. H. Whitfield, 2014: Application potential of four nontraditional similarity metrics in hydrometeorology. *Journal of Hydrometeorology*, **15 (5)**, 1862–1880.

Murphy, A. H., 1995: The coefficients of correlation and determination as measures of performance in forecast verification. *Weather and Forecasting*, **10 (4)**, 681–688.

Romero, R., C. Doswell III, and R. Riosalido, 2001: Observations and fine-grid simulations of a convective outbreak in northeastern Spain: Importance of diurnal forcing and convective cold pools. *Monthly Weather Review*, **129 (9)**, 2157–2182.

Speer, M. S., and L. M. Leslie, 2002: The prediction of two cases of severe convection: implications for forecast guidance. *Meteorological Atmospheric Physics*, **80 (1-4)**, 165–175.

Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: *Survey of common verification methods in meteorology*. World Meteorological Organization Geneva.

Swinbank, R., and Coauthors, 2016: The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*, **97 (1)**, 49–67.

Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *ECMWF Workshop on Predictability*, Vol. 1, 25.

Wang, Z., and A. C. Bovik, 2002: A universal image quality index. *IEEE Signal Processing Letters*, **9 (3)**, 81–84, doi:10.1109/97.995823, URL http://dx.doi.org/10.1109/97.995823.

Wang, Z., and A. C. Bovik, 2009: Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, **26**, 98–117.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, 2004: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Pro-*

cessing, **13**, 600–612.

Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review*, **125 (4)**, 527–548.

Wernli, H., M. Paulat, M. Hagen, and C. Frei, 2008: SAL-A novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, **136**, 4470–4487.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, Vol. 100. Academic Press.